



Impact Evaluation of Math Professional Development

OMB Clearance Request, Part B

PREPARED BY:

American Institutes for Research[®]
1000 Thomas Jefferson Street, NW, Suite 200
Washington, DC 20007-3835

PREPARED FOR:

U.S. Department of Education
Institute for Education Sciences

December 10, 2012

Contents

	Page
Description of the Impact Evaluation of Math Professional Development.....	2
B. Description of Statistical Methods.....	11
1. Respondent Universe and Sampling Methods.....	11
2. Procedures for Data Collection.....	17
3. Procedures to Maximize Response Rates.....	21
4. Pilot-Testing Instruments.....	22
5. Names of Statistical and Methodological Consultants and Data Collectors.....	22
References.....	24

List of Exhibits

- Exhibit 1. Respondent Universe for Recruitment
- Exhibit 2. Respondent Universe for Proposed Recruitment Activities
- Exhibit 3. Mapping of Recruitment Screening Protocol Items to Constructs
- Exhibit 4. Respondent Universe for Proposed Data Collection Activities
- Exhibit 5. MDES for Main Outcome Measures
- Exhibit 6. Alignment of Intel Math Content and the Teacher Knowledge Item Pool (MTEL)
- Exhibit 7. Mapping of End-of-Year Teacher Survey Items to Constructs
- Exhibit 8. Mapping of Requested Extant Data to Constructs
- Exhibit 9. Teacher Response Rates from Selected Prior AIR Studies

List of Appendices

- Appendix A – Recruitment Materials
 - A-1 IES Letter of Support
 - A-2 District Screening Protocol
 - A-3 School Screening Form
 - A-4 Teacher Interest Form
- Appendix B – Data Collection Instruments
 - B-1 Teacher End-of-Year Survey
 - B-2 Extant Data Collection Protocol

Study Background

The need to improve U.S. students' math achievement is clear. A minority of U.S. students score at or above proficient levels in math and science on the National Assessment of Educational Progress (NAEP), and recent math and science achievement scores on the Programme for International Student Assessment consistently place U.S. 15-year-old students no higher than average internationally (National Center for Education Statistics, 2011; National Mathematics Advisory Panel; 2008; National Science Board, 2012). Teacher professional development (PD) is considered to be an important pathway to improving teaching and learning in general, and improving mathematics teaching and learning, in particular, and federal, state and local governments invest billions of dollars each year to support the development and delivery of preservice and inservice training.

Despite these investments in PD, there is limited rigorous evidence of the effectiveness of specific PD strategies. In particular, there is a lack of evidence about PD that places a strong emphasis on boosting elementary teachers' content knowledge and transferring that knowledge to the classroom, though mathematicians and math educators have argued that gaps in elementary teachers' math content knowledge must be addressed. Recognizing this need, the National Center for Education Evaluation (NCEE) at the Institute of Education Sciences (IES) commissioned a study to evaluate the impact of an intensive, content-focused PD program on teachers' content knowledge, classroom practice and student achievement. This study will contribute much-needed information and evidence to a field in need of high quality information about improving students' math performance and teacher quality in our nation's schools.

The Impact Evaluation of Math Professional Development is designed to examine the implementation and impact of a widely-used, intensive PD program that has a strong emphasis on developing teachers' content knowledge and supporting the transfer of knowledge into the classroom. The PD program was determined by the U.S. Department of Education to be the most promising, scalable intervention with these features. More specifically, the program being tested in this evaluation includes (1) the Intel Math Program, an 80-hour course to be delivered in summer/early fall 2013, (2) the Math Learning Community (MLC), a 10-hour follow-up component in which groups of teachers collaboratively analyze student work on topics covered in the summer Intel Math course, and (3) a 3-hour video feedback component, in which teachers receive feedback regarding the quality and clarity of their mathematical explanations from video lesson excerpts on topics emphasized in Intel Math and the MLCs. All of these activities will be delivered by trained Intel course instructors and MLC facilitators beginning in summer 2013. By testing an intervention that incorporates features the available research suggests are essential, this study has high policy value and relevance to the field.

To determine the impact of the PD program on teacher knowledge, teacher practice and student achievement, a purposive sample of six districts and approximately 200 4th grade teachers will be recruited to participate in the study in 2013-14. These districts will be selected according to multiple criteria, including size, structure of math instruction, student composition of math classes, content of math instruction, and prevalence of competing curricular or PD initiatives occurring during the 2013-14 school year.

Recruitment in eligible districts will focus primarily on identifying teachers in 4th grade who are willing to participate in the study and who have the support of their principals to do so. The study focus is on upper elementary teachers because they are less likely to have a strong math background than middle school teachers. The particular focus on one grade level for the evaluation is for the sake of clarity, simplicity, and to control study costs. Fourth grade was chosen over 5th grade because it falls right at the center of the K-8 spectrum covered in the Intel Math course and the Intel topics are more closely aligned with topics typically covered in 4th grade than those in 5th grade.

The evaluation design for the Impact Evaluation of Math Professional Development involves the random assignment of approximately 200 volunteer grade 4 teachers in six districts to one of two conditions: Treatment teachers will be offered the study's PD intervention from summer 2013 through spring 2014; control teachers will be offered their district's business-as-usual PD. The teacher-level random assignment will be conducted within-school and within-fourth-grade to maximize the study's statistical power to detect impacts. In cases where there is an even number of fourth-grade teachers, the treatment and control groups will be of equal size. In cases where there is an odd number of fourth-grade teachers, the treatment group will always exceed the control group by one teacher. Providing the extra treatment teacher in these cases will help offset potentially greater attrition from the treatment group (which must participate in the PD intervention), as well as increase the likelihood of having multiple treatment teachers from the same school, which is preferable from the PD vendors' perspective. During the one-year implementation period, data will be collected to support analyses of the implementation and impact of the PD program, with final data collection of teacher and student outcomes in spring 2014 and all analyses and reporting completed by winter 2016.

NCEE is requesting clearance to carry out recruitment and data collection activities for the Impact Evaluation of Math Professional Development. Recruitment activities include contacting a purposive sample of districts, schools and teachers to establish their eligibility and interest in participating in the study. Data collection activities include administering three teacher knowledge assessments (baseline and two follow-ups), an end-of-year teacher survey, an end-of-year student assessment, and an extant data collection protocol.

This evaluation is authorized by Title IX, Part F of the Elementary and Secondary Education Act, section 9601 as amended by the "No Child Left Behind Act of 2001" (20 USC 7941).

Research Questions

The study is designed to answer two main research questions, the first focusing on the impact of the PD program on teacher and student outcomes and the second focusing on program implementation.

- RQ1.** What is the average impact on teachers' content knowledge, teachers' classroom practices, and student achievement, of offering a specialized PD intervention relative to "business as usual" PD?
- RQ2.** How is the PD intervention implemented? What challenges were encountered during the process of implementing the intervention?

The study's main outcome measures are teachers' content knowledge, classroom practices, and student achievement. We will measure teachers' content knowledge at three timepoints: first, at baseline (summer 2013), second after completion of the content-intensive summer PD component (fall 2013), and third at the end of the school year (June 2014). Content knowledge will be measured with a mathematics assessment composed of items in the Massachusetts Test for Educator Licensure (MTEL) for elementary math teachers. We will measure teachers' classroom practice in spring 2014 using the Mathematical Quality of Instruction (MQI) instrument, which is a previously-validated observation protocol applied to video-recorded observations of teachers' lessons. Finally, we will measure student achievement at the end of the 2013-14 school year with two instruments: (1) the state math assessment (standardized across states) at baseline (spring 2013) and follow-up (spring 2014), and (2) a study-administered assessment in a random sample of 10 students per participating teacher (total N = 2,000). In addition, the study team will administer a survey in June 2014 to collect teacher background characteristics for use as covariates in the impact analyses for RQ1 and information about implementation and PD service contrast for RQ2.

Intervention

The existing literature suggests that any high-quality PD program should have: (1) a heavy emphasis on comprehensively developing mathematical content knowledge and (2) a well-defined teacher support structure to ensure that the training is transferred into the classroom. Therefore, IES is interested in testing a PD program that has an intensive and comprehensive mathematical content component (i.e., an intensive summer institute for math teachers) and supports teacher efforts to incorporate such learning into their everyday teaching (e.g., structured professional learning communities that reinforce the implementation of PD practices). The math content in PD programs is typically taught using one of two general approaches. The first approach is more similar to a traditional university mathematics course, where the focus is directly on teaching teachers the pure math content underlying the topics to be taught in the classroom and on enabling teachers to actually practice doing the mathematics. The second approach typically involves the use of analysis of actual student work and classroom case scenarios to indirectly strengthen teacher participants' own understanding of the underlying math content. Given that prior studies such as Garet et al (2011) have primarily tested PD programs employing the second approach, IES is instead interested in testing a PD program employing a more explicit approach to teaching math content. IES is also interested in testing a PD program that is presently policy relevant, meaning that the PD is currently being implemented across multiple districts and states, and could be similarly implemented consistently in a large-scale evaluation or by other sites if desired. Practically, this implies that the PD is an "off-the-shelf" program that requires no customization and possesses the infrastructure for scale-up across multiple states.

After examining several existing math PD programs (including Developing Mathematical Ideas, Lesson Study with Fractions Toolkit, and Math Solutions), IES identified the Intel Math Program in combination with a Mathematics Learning Community (MLC) as a comprehensive and intensive set of PD activities that meet the requirements described above. Since its inception in 2006, over \$5 million has been invested in Intel Math while being implemented in at least seven states and featured in federally-funded Math Science Partnership grants for Arizona and

Massachusetts, as well as in Massachusetts' Race to the Top agenda. Although a few exploratory studies have been conducted, there is no rigorous evidence of the impacts for this type of program on teacher and student outcomes, despite its growing popularity.

Thus, the intervention being tested in this study is designed to support teachers' development of math content knowledge and the transfer of that knowledge to students. The intervention has three components, the core of which is the Intel Math Program. Intel Math is an 80-hour, university course-like program that focuses on strengthening teachers' mathematics content knowledge; it will be delivered primarily in summer 2013. The PD intervention continues into the 2013-14 school year with a support structure designed for use with Intel Math, the Mathematics Learning Community (MLC). The MLC will provide 10-hours of follow-up, collaborative meetings (five, two-hour meetings) that focus on analyzing student work on topics addressed in Intel Math. The MLC facilitators will also deliver direct feedback to participating teachers on their classroom practice (video recorded three times during the school year) with a focus on the quality and clarity of the teachers' mathematical explanations on topics addressed in Intel Math and the MLC meetings. Teachers will spend three hours across the three video feedback cycles, bringing the total of the three-part intervention to 93 hours.

The focus of the evaluation is on the effects of the intervention on 4th grade teachers and their students. However, teachers of other grade levels will be invited to participate in parts of the intervention, as described in the following sections.

Intel Math

Intel Math is a widely used PD program. The program is currently being implemented in 11 states and 49 cohorts of teachers – more than 1,000 K-8 teachers in total. The program has a strong focus on improving teachers' math content knowledge (Mundry et al., 2011). About 90 percent of the focus is on foundational math content for K–8 teachers; the other 10 percent is on pedagogy. Teachers learn the content primarily by solving conceptual and computational math problems grounded in real-world settings, and receive feedback from their instructors (each course is co-taught by a university mathematician and mathematics educator). Teachers are encouraged to use and share multiple solution methods, and the course emphasizes helping teachers see how arithmetic and algebra are interconnected and represent the same mathematical ideas. The 10 percent of the course that is devoted to pedagogy examines strategies associated with teaching the content in each unit, mostly through the examination of student work samples.

The topics of the Intel Math course are as follows:

- Unit 1: addition
- Unit 2: subtraction
- Unit 3: multiplication
- Unit 4: division
- Unit 5: operations with fractions
- Unit 6: rational numbers
- Unit 7: linear relations
- Unit 8: functions

Among the eight units in the course, Units 1–5 focus directly on topics included in the Grade 4 Common Core State Standards in Mathematics (CCSSM) (addition, subtraction, multiplication,

division, and meaning of fractions). Units 6–8 focus on ratio/proportion, algebra, and linear functions, which are topics important for 4th grade teachers to know so that they can provide instruction that appropriately lays the foundation for students’ learning in future grades. Intel Math is delivered face-to-face, rather than remotely, because of the emphasis on problem-solving, solution-sharing and cooperative learning.

The course is typically taught to teachers of multiple grade levels, which allows for discussions about how concepts develop over time and opportunities for teachers to deepen their understanding of the math that come before and after the math that they teach. In typical implementation of Intel, schools are also encouraged to ensure that at least two teachers participate together; this gives teachers partners for transportation and greater opportunity to continuing discussing course content.

In order to implement Intel Math in the typical manner for the study, a mix of teachers in grades K-8 will be invited to participate along with the 4th grade study teachers. We will recruit approximately 10 additional teachers—five teachers from grades K-3 and five from grades 5-8 to provide a balance of teachers in grades below and above the targeted grade 4. The teachers in other grade levels will be selected in an effort to ensure that all or most of the 4th grade study teachers who are randomized to the treatment group have another teacher from their school participating in the PD intervention.

The study’s main focus is on improving teachers’ content knowledge, and the Intel Math program is the core of the study’s PD intervention. The other two components, described next, are intended to support the enactment of teachers’ content knowledge in their classroom practice.

Mathematics Learning Community (MLC)

The MLC offers a support structure to help teachers transfer the content they are learning through Intel Math to their students’ work. The centerpiece of each two-hour session is the analysis of student work samples using a standardized protocol. Looking at student work encourages teachers to think about the underlying math concepts in problems with which students struggle, by analyzing different student approaches, solutions, common errors, and misconceptions. According to the MLC developers, the learning communities function best when they are implemented primarily in the fall, which is relatively close to when teachers studied the same topics in the Intel summer course; strongly supported by district and school leadership; and integrated into the district’s instructional system—curriculum, pacing guides, curriculum, assessments—as opposed to being an add-on to the full set of instructional and assessment demands facing teachers on a regular basis. They are typically but not always implemented with teachers from multiple grade levels; for example, teachers in the 3-5 grade band.

The complete MLC program includes 15 sessions that are typically implemented over two years and focus on topics spanning grades K-8. However, with the study’s focus on 4th-grade teachers and duration of one year, we have selected 5 MLC sessions that are most aligned to grade 4 topics and Units 1–6 of the Intel course, maximizing the coherence of the PD intervention within the instructional context of each district.

The participants in the MLCs will include all of the 4th grade study teachers and those teachers in grades 3 and 5 who participated in the Intel course. This will ensure that the configuration of

the MLCs for the study is similar to that in typical implementations of the program, where including a mix of teachers from the 3-5 grade band is viewed as desirable. As noted, high priority for selection will be given to 3rd and 5th grade teachers in schools where only one 4th grade study teacher will be randomized to the treatment group (because, e.g. there are only two volunteer 4th grade teachers in the school).

Video Feedback

The third component of the intervention includes three video feedback cycles that are designed to help teachers improve the quality, clarity and coherence of their explanations of topics that are central to grade 4, as defined by Intel Math, and as the topics appear in each district’s pacing guide (multiplication, division and fraction concepts). To plan each cycle, the MLC facilitator will work with each teacher to select an appropriate lesson to be videotaped, support the teacher in video recording the lesson, and send the video to be scored by trained coders at Harvard University. The Harvard coders will score each lesson using the MQI, which rates the quality and clarity of the mathematical explanations and discourse in the lesson. The MLC facilitator will then use the MQI scores and illustrative clips provided by the Harvard coders to prepare feedback for the teachers. The feedback will be discussed in one-hour, one-on-one meetings with the teachers; the meeting time will include making a plan for improving the quality of explanations on these topics as they are revisited in future lessons and as they relate to other topics that will be introduced later in the year. (For example, if the focus of an initial lesson is clarifying the meaning of numerator and denominator, the clarity of language and presentation of these concepts would be revisited when students add and subtract fractions later in the year).

The participants for the video feedback component are the 4th grade teachers in the study sample.

Together, the three intervention components are intended to boost teachers’ content knowledge and provide a support structure for transferring that knowledge to the classroom.

Analytic Strategy for Analyses of Program Impacts

RQ1 assesses the impact of the PD intervention on teacher and student outcomes. Our analytic strategy for both sets of outcomes are described next.

Intent-to-Treat Impact Analyses. The main analyses testing the effects of the PD intervention on teacher and student outcomes (RQ1) take an intent-to-treat (ITT) approach, meaning that all teachers who were randomly assigned to the treatment and control groups are included in the analysis sample. Teacher knowledge will be based on the following regression model:

$$Y_{jk} = \sum_{s=1}^S \beta_{0k} * (SCHOOL_s)_{jk} + \sum_{d=1}^6 \beta_{1d} * PD_{jk} * (DISTRICT_d)_{jk} + \beta_3 * W_{jk} + r_{jk} \quad (1)$$

where Y_{jk} is a measure of teacher knowledge for teacher j in school k , $SCHOOL_s$ is a set of indicators for the S study schools, PD_{jk} is an indicator for treatment status of teacher j in school k , $DISTRICT_d$ is a set of indicators for the six school districts, and W_{jk} is a vector of teacher

background characteristics for teacher j in school k (e.g., the baseline measure of the outcome). β_{0k} represents the average outcome among control teachers in school k , and β_{1d} captures the treatment effect in district d . The overall treatment effect across all six school districts can be computed as a weighted average, with each school district weighted by the number of treatment teachers in the school district. Thus, the overall treatment effect represents the effect of the PD program on a typical treatment teacher in the sample.

For the analyses of treatment effects on teachers classroom practices, we will extend Equation 1 to a two-level hierarchical linear model (HLM) model to explicitly take into account the clustering of lessons observed within teachers. The model will be specified as follows:

Level 1 (lessons):

$$Y_{ijk} = \pi_{0jk} + \varepsilon_{ijk} \quad (2)$$

where Y_{ijk} is the MQI rating of lesson i taught by teacher j in school k , π_{0jk} is the average rating of the lessons observed for teacher j in school k , and ε_{ijk} is a random error associated with a given lesson.

Level 2 (teachers):

$$\pi_{0jk} = \sum_{s=1}^S \beta_{00k} * (SCHOO L_s)_{jk} + \sum_{d=1}^6 \beta_{01d} * PD_{jk} * (DISTRICT_d)_{jk} + \beta_{03} * W_{jk} + r_{0jk} \quad (3)$$

The interpretation of Equation 3 is similar to that for Equation 1. In particular, β_{01d} represents the treatment effect on the average lesson rating for individual teachers in district d , and the overall treatment effect can be computed as a weighted average effect across the six school districts. To test the PD program's effect on student achievement at the end of the program year (spring 2014), we will construct the following model where students are nested within teachers:

Level 1 (students):

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk} * X_{ijk} + \varepsilon_{ijk} \quad (4)$$

where Y_{ijk} is the test score of student i taught by teacher j in school k , and X_{ijk} is a vector of demographic characteristics and prior year achievement score of student i taught by teacher j in school k , grand-mean centered. The intercept equation at the school level (Equation 5 below) is identical to Equation 3, with similar interpretations of the terms. The student-level covariate slopes are fixed to their grand means at the school level (Equation 6).

Level 2 (teachers):

$$\pi_{0jk} = \sum_{s=1}^S \beta_{00k} * (SCHOO L_s)_{jk} + \sum_{d=1}^6 \beta_{01d} * PD_{jk} * (DISTRICT_d)_{jk} + \beta_{03} * W_{jk} + r_{0jk} \quad (5)$$

$$\pi_{1jk} = \beta_{10k} \quad (6)$$

Treatment-on-the-Treated Analyses. It is possible that some teachers assigned to the treatment may not attend the PD activities. Although ITT analyses provide valid estimates of the treatment effects on teachers assigned to the PD program, they may underestimate the treatment effects on teachers who actually attended the PD activities if the number of no shows is not trivial. In that case, we will supplement the ITT analyses with treatment-on-the-treated (TOT) analyses to assess the treatment effects on teachers induced to participate by treatment assignment.

We will conduct the TOT analyses using a standard instrumental variable (IV) approach, where the treatment assignment will serve as the instrument for PD participation (Angrist, Imbens, & Rubin, 1996; Gennetian, Morris, Bos, & Bloom, 2005). During the first stage of the IV analysis, treatment assignment (i.e., IV) is used to obtain the predicted probabilities of PD participation. The predicted values, instead of the original values, of PD participation are then used in the second stage to predict the outcome. The resulting IV estimate of the effect of PD participation can be interpreted as the treatment effect on treatment teachers who were induced to fully participate in the PD program because of treatment assignment (i.e., the local average treatment effect or the treatment effect on compliers).

Dosage Analyses. Given that the level of participation in PD activities (i.e., dosage) is likely to vary across treatment teachers, we will conduct dosage analyses to examine the extent to which the level of PD participation is related to the size of the treatment effect. Our proposed design where teachers are randomly assigned within schools lends itself to a dosage analysis based on the following HLM model, using teacher knowledge as an illustration:

Level 1 (teachers):

$$Y_{jk} = \beta_{0k} + \beta_{1k} * PD_{jk} + \beta_{2k} * W_{jk} + r_{jk} \quad (7)$$

Level 2 (schools):

$$\beta_{0k} = \sum_{d=1}^6 \gamma_{00d} * (DISTRICT_d)_k + \gamma_{01} * DOSAGE_k + u_{0k} \quad (8)$$

$$\beta_{1k} = \sum_{d=1}^6 \gamma_{10d} * (DISTRICT_d)_k + \gamma_{11} * DOSAGE_k + u_{1k} \quad (9)$$

$$\beta_{2k} = \gamma_{20} \quad (10)$$

In the level 2 model, *DOSAGE* is a school-level measure of PD participation (e.g., the total number of hours of math PD as part of the study intervention) computed as the average participation level among treatment teachers within a given school.¹ The parameter of primary

¹ Because dosage is observed only for treatment teachers, it cannot be used as a teacher-level predictor. Therefore, we use a school-level dosage measure to predict the school-specific treatment effect. Because dosage is measured at the school level, it cannot be used in a model with school fixed effects (there is no variation in school-level dosage within schools); therefore, we treat schools as random effects in this analysis. Before running the fully specified HLM model, we will first estimate the

interest from this dosage analysis is γ_{11} , which indicates the extent to which the treatment effect is larger in schools where the average level of participation among treatment teachers is higher, controlling for whether there is only one or multiple treatment teachers in a school.² Similar analyses could be conducted to estimate the relationship between dosage and teacher practice or student achievement, using a three-level model.

Analytic Strategy for Analyses of Program Implementation

To describe how the PD intervention was implemented and the challenges associated with implementation (RQ2), we will conduct descriptive analyses of data collected using the measures previously described (with expanded descriptions in the following section). These analyses will describe (1) the extent to which the PD intervention (Intel and MLC) was delivered as intended (fidelity); (2) the proportion of the intended number of hours of the intervention that treatment teachers received (participation); and (3) the difference between PD received by treatment and control teachers (service contrast). Our fidelity of the Intel course and MLC meetings and service contrast analyses will describe the duration, content emphasis, coverage of planned materials, type of learning activities and active engagement of the participants.

We will use the information provided in the teacher surveys to describe the treatment contrast in the number of hours and types of study-relevant PD activities in which teachers participated during the 2013-14 school year. Study-relevant PD activities include extended math content-focused workshops (1/2 day or longer), collaborative meetings that focus on analyzing student work or data (e.g., lesson study) and opportunities for teachers to receive feedback on the quality of their mathematical explanations, through videotaped lessons or direct observations.

between-school variance of the treatment effect based on a model without any school-level predictors. A meaningful dosage analysis is warranted only if there is significant between-school variation in the treatment effect.

² A similar model could be estimated with a variable at level 2 indicating whether a school has one or more than one treatment teacher, to test whether the treatment effect is larger if multiple teachers in a school are assigned to participate in the PD program.

Supporting Statement for Paperwork Reduction Act Submission

Supporting Statement Part B below addresses the following sample recruitment and data collection aspects of the study: respondent universe and sampling, procedures for data collection, procedures to maximize response rates, pilot-testing instruments, and names of statistical and methodological consultants and data collectors.

B. Description of Statistical Methods

1. Respondent Universe and Sampling Methods

In this section, we describe the respondent universe and sampling methods for recruitment activities and data collection activities.

Study Recruitment Activities

The Impact Evaluation of Math Professional Development will test the effectiveness of an approach to content-intensive PD for 4th grade teachers. It will not employ random sampling of districts, schools or teachers for the purpose of generalization. Instead, districts will be screened and recruited based on characteristics required by the study design. For example, the study design seeks districts that do not already widely use Intel Math or have similarly content-focused PD planned for school year 2013-2014 for their 4th grade teachers, and have at least 16 elementary schools with at least two 4th grade math teachers.

Similarly, schools within districts will be recruited based on the requirements of the study design. For example, each school must have non-departmentalized math instruction, at least two 4th grade math teachers, and not sort students by ability level into classes for math instruction. At the teacher level, due to the time commitment required for the study-provided math PD, only teachers that volunteer will be considered eligible for study participation. To achieve a study sample of 200 volunteer teachers in approximately 11 schools within each of 6 districts, the study team will conduct district and school/teacher level recruitment activities.

Exhibit 1 describes the steps and associated respondent universes for study recruitment. This process includes pre-screening at the district level and interviews at the district, school and teacher levels. Each of these steps in the recruitment process is described more fully in the following sections: (1) identifying the pool of districts to be screened, (2) conducting district-level screening interviews, (3) prioritizing districts for recruitment, (4) recruiting eligible districts, (5) recruiting eligible schools and 4th grade math teachers, and (6) negotiating final agreements.

Exhibit 1. Respondent Universe for Recruitment

Steps in the Recruitment Process	Respondent Universe
District level prescreening	Districts located in all states
District screening interviews	80 most qualifying districts (30 highest qualifying after interviews)
First district visit	15 of the 30 highest qualifying districts that are interested in study participation
Second district visit, including visits to interested schools	8 of the 15 interested districts, approximately 114 school administrators and 340 teachers

Identifying the Pool of Districts to Be Screened. To identify the pool of districts, we will use the Common Core of Data to identify states that have at least one district with 16 or more elementary schools containing at least two grade 4 teachers. Having districts of this size is necessary to satisfy the study design, which assumes 16-18 grade 4 teachers from each district will participate in the study PD. Only states with at least one district of this size will be screened for potential districts. We will also collect information from these states' Websites about whether major initiatives (e.g., new state curriculum, evaluation or assessment initiative) will be occurring during the 2013-14 school year. States will not be excluded based on these criteria, but the information will be used in subsequent communications with districts—i.e., to acknowledge that the study team is aware of the initiative and to gauge the extent to which the initiative might be disruptive in terms of implementing the study during summer 2013 and the 2013-14 school year. Among these districts—we anticipate 400-500 districts will satisfy the size requirement—we will prioritize the 80 districts to be screened based on two criteria: districts we have worked with or have knowledge of that might be interested in participating in the study and districts that are located in different states and different geographic regions of the country.

Conducting District-Level Screening Interviews. After the pool of 80 districts has been selected based on the previously mentioned criteria, the study team will send an informational e-mail to each of these districts that includes a letter from IES introducing the study (see Appendix A: A-1). After the e-mail is sent, a recruitment team member will call each district to inform the district about the study and ask the district to participate in a telephone call, which will be guided by the district-level screening protocol presented in Appendix A: A-2 and is described in the section Respondent Universe and Screening Materials for Recruitment.

A district will be determined to be eligible for the study after the study team confirms through the phone call that the district meets the following criteria:

- *Size:* Districts must have 16 or more elementary schools with at least two 4th grade teachers per school who may be interested in volunteering to participate in the study);
- *Structure of math instruction – non-departmentalized and not ability tracked:* Districts must have schools in which (a) teachers are non-departmentalized but rather they teach all or most subjects including math, and (b) students are not sorted by ability level into classes;
- *Content of math instruction – no major changes:* Districts must indicate that the curricula they plan to implement in 2013-14 is not a major change or overhaul from the prior year;
- *Other PD activities or initiatives:* Districts must indicate that they do not plan to (a) provide 4th grade math teachers PD similar in focus or intensity to what will be provided by the study during the school year 2013-2014 or (b) initiate a district-wide initiative that

might interfere with grade 4 teachers willingness to participate or ability to benefit from the PD provided by the study.

The screening protocol is designed to allow early termination of the interview if a district does not meet these criteria. We anticipate that approximately 30 of the 80 districts will meet the aforementioned criteria.

Prioritizing Districts for Recruitment. Among the 30 districts that we expect will pass the initial screen, some will be more appropriate candidates for the study than others. AIR staff will use the additional information gathered in the prescreening and screening interviews to prioritize eligible districts for recruitment efforts. The following criteria will be used:

- *Interest:* Districts that signal greater interest will be given higher priority. Interviewers usually receive some signals about the district's level of interest in the study even though the screening protocol contains no questions about interest. Whenever possible, we will also incorporate our prior work and knowledge of specific districts to gauge the interest level of districts.
- *Feasibility of implementation:* Districts that have fewer competing initiatives for 4th grade teachers, whether in math or in other subjects, will be given higher priority. In addition, districts that have consistent use of curriculum program across schools will be given priority.
- *Geographic location:* As with states, some districts may receive higher priority due to geographic proximity to where the PD providers already have instructors and trainers available and if the location of the district expands the geographic diversity of the potential study sample.

Considering these criteria while canvassing the 30 eligible districts will allow AIR to prioritize the districts before initiating requests for site visits.

Recruiting Eligible Districts. Past experience indicates that site visits to districts are necessary to ensure eligibility and reach final agreement on participation. Senior recruitment team members will follow up by telephone with the highest-priority districts identified through the screening process to determine which district officials must be involved in making the decision about participation, communicate the specific benefits of participating in the study, describe the ways in which the study will minimize the burden of participants, and determine whether the district is sufficiently interested in the study and, if so, offer to visit the district to further discuss participation.

We anticipate 15 of the 30 districts to express sufficient interest because of the benefits of participation. Senior staff from the study team will visit these 15 interested districts. The site visits will allow us to present the study in-person, discuss the benefits and responsibilities of participation with additional district officials and principals, and respond to any questions and concerns that they might have.

During the recruitment visit, study staff will work with the district to identify schools that meet the following criteria:

- *Size:* Schools must have at least two 4th grade teachers who are likely to be interested in volunteering to participate in the study;
- *Structure of math instruction – non-departmentalized and not ability tracked:* Schools must have schools in which (a) teachers are non-departmentalized but rather they teach all or most subjects including math, and (b) students are not sorted by ability level into classes;
- *Content of math instruction:* Schools must indicate that the curricula they plan to implement in 2013-14 is not a major change or overhaul from the prior year;
- *Other PD activities or school-based initiatives:* Schools must indicate that they do not plan to (a) provide 4th grade math teachers PD similar in focus or intensity to what will be provided by the study during the school year 2013-2014 or (b) initiate a school-wide initiative that might interfere with grade 4 teachers willingness to participate or ability to benefit from the PD provided by the study;
- *Interest:* Schools that signal greater interest in the study will be given higher priority.

To maximize the number of schools and teachers to volunteer to participate, AIR will ask district leadership to convey support for the study and to set the expectation that the qualifying schools will consider participation in the study.

Recruiting Eligible Schools and 4th Grade Math Teachers. We expect eight districts will remain interested in participating in the study after the first district recruitment visit, and we will schedule a second visit to these districts. The primary goals of the second visit are to (1) confirm school eligibility criteria received from the district, and (2) recruit schools and 4th grade math teachers to the study. The recruitment team members, with the input from district personnel, will either organize meetings for clusters of schools or will visit each eligible and potentially interested campus separately.

A school's eligibility to participate in the study will be reconfirmed in a meeting with the principal by using the criteria listed in the previous section (size, structure of math instruction, content of math instruction, other PD activities or school-based initiatives, and interest). In addition, the school visits provide an opportunity to inform teachers about the study and recruit potential teachers in schools that meet the criteria, especially those teachers who express interest in participating in this type of PD program and have the support of their principals to do so.

Study staff have identified several benefits to participation as part of the recruitment process. First, the U.S. Department of Education invested resources to determine that the study PD program (Intel Math and MLC) is a high-quality, scalable program designed to improve teachers' mathematical content knowledge and connections to the classroom. In addition, districts will not be charged for the PD which is provided as the study's intervention, and teachers in the treatment group will receive the PD for free. The study also pays for the salary of two MLC facilitators selected from each district—these costs include travel and labor associated with participating in the MLC facilitator and MQI trainings and labor associated with implementing the MLC and video feedback PD activities during the 2013-14 school year. For the video feedback component of the study, each district will also receive the requisite hardware and software for use during the 2013-14 school year. As an additional study incentive, we plan to request permission from the

U.S. Department of Education to allow participating districts to keep the video equipment at the conclusion of the study.

Negotiating Final Agreements. Shortly after the second district visit, district administrators will be asked to reach a final agreement to participate, and a memorandum of understanding with interested districts will be prepared. As a condition of participation, districts also must ask the principals of eligible and willing elementary schools and the volunteer 4th grade teachers within these schools) to submit signed statements reflecting an intention to participate. If necessary, project staff will make additional phone calls or visits to build consensus and obtain commitment from principals and teachers. The principal and teacher signatures are expected to be gathered shortly after the district memorandum of understanding is obtained in order to allow random assignment of teachers within schools before the start of the 2013–14 school year.

Respondent Universe and Screening Materials for Recruitment

As previously discussed, recruitment activities will target districts, schools and teachers. Exhibit 2 outlines the respondent universe for these recruitment activities and the protocols that will be used to gauge the interest of potential study participants. After factoring in anticipated response rates, we expect to involve roughly 68 districts, 114 school administrators and 340 teachers in recruitment outreach activities, ranging from phone calls to in-person site visits.

Exhibit 2. Respondent Universe for Proposed Recruitment Activities

Data Source	Respondent Universe (Administrators & Teachers)
District Screening Protocol	68
School Screening Protocol	114
Teacher Interest Form	340

The three screening protocols that will be used to recruit districts, schools and teachers are included in this package in Appendixes A: A-2, A-3 and A-4. The items on the screening protocols are mapped to the constructs that they measure in Exhibit 3. The district screening protocol will be administered by project staff via telephone calls with district personnel and the school screening protocol and teacher interest forms will be administered in-person by research staff during site visits.

Exhibit 3. Mapping of Recruitment Screening Protocol Items to Constructs

District Screening Protocol	
Constructs	Items
Current math instruction	1-7
Planned changes in district practices affecting 4th grade instruction	8-9
Student testing	10-12
Professional development for 4th grade math teachers	13-19
School Screening Protocol	
Instructional format for mathematics instruction	1-3
Planned changes in school affecting 4th grade instruction	4
Professional development and curriculum for 4th grade math	5-10
Teacher Interest Form	
Eligibility for study (teaching at target grade level, non-departmentalized, mixed-ability class)	1-3
Interest and availability to participate in study	4-5

Study Data Collection Activities

Following completion of recruitment and the launch of the evaluation, we expect to collect the types of data from the respondent universe presented in Exhibit 4. For all teacher-level data collection activities, the respondent universe will be all 200 teachers participating in the study; the only exception being data collection regarding implementation of the PD intervention, which will be conducted with only the 100 teachers randomly assigned to the treatment group. For the study-administered student assessment, the respondent universe will be 10 randomly selected students in each teacher’s classroom. The student-level extant data collection respondent universe will be all students in participating teachers’ classrooms, estimated to be on average 20 students per classroom.

Exhibit 4. Respondent Universe for Proposed Data Collection Activities

Data Source	Number of Records		Collection Schedule			
	Treatment	Control	Summer 2013	Fall 2013	Winter 2014	Spring 2014
Teacher knowledge test (<i>N</i> teachers)	100	100	X	X		X
Teacher survey (<i>N</i> teachers)	100	100				X
Video observations for evaluation (<i>N</i> observations)	300	300		X		X
Study administered student test (<i>N</i> students)	1,000	1,000				X
District archival records; approximate <i>N</i> students plus teachers per condition)	2,100	2,100	X			X
Fidelity and log data on PD intervention (Intel, MLC, and Video Feedback)	100	(N/A)	X	X	X	X

*Bolded items are data collection instruments involving burden of study participants.

To assess the statistical power of the study design, we draw on recent literature on power analysis for group randomized trials (Schochet, 2008; Spybrook, Raudenbush, Congdon, & Martinez, 2009) to calculate the variance components and estimate the minimum detectable effect sizes (MDESs) for student achievement outcomes, teacher knowledge and practice outcomes, and student achievement. We derive assumptions from prior studies about the proportion of the variance in the outcome measures that are between schools and between teachers within schools, the percentage of outcome variance explained by covariates, the number of districts and the number of schools per district, the number of teachers per school, the number of students per teacher, and the number of teachers observed per school. To reflect both optimistic and cautious assumptions, we have calculated MDES ranges for our main outcome measures (Exhibit 5).

Exhibit 5. MDES for Main Outcome Measures

Outcome Measure	MDES
Teacher knowledge	0.20 – 0.27
Classroom practice	0.28 – 0.39
Student achievement	0.08 – 0.12

2. Procedures for Data Collection

AIR project staff will manage data collection to ensure quality and timeliness. The data collection includes a baseline, mid-year and end-of-year teacher knowledge assessment, an end-of-year teacher survey, an end-of-year student assessment, and an extant data collection protocol (also summarized in Part A of this submission). The teacher knowledge assessments and teacher survey will be administered in person on paper to all 200 teachers participating in the study. The archival record requests will be sent via e-mail to each study district and followed up via phone calls. The timeline for different data collection activities is as follows:

Data Collection Timeline

1. *Summer 2013*. Baseline teacher knowledge assessment administered by study staff.
2. *Summer 2013*. Extant records for all students in participating teachers' classes as of start of 2013-14 school year (baseline).
3. *September-October 2013*. Fall post-Intel teacher knowledge assessment administered by study staff.
4. *October 2013/March 2014*. Teacher video observations (treatment and control teachers).
5. *March 2014*. Extant records for all students in participating teachers' classes as of spring 2014.
6. *May 2014*. Study-administered student assessment.
7. *June 2014*. Follow-up teacher knowledge assessment and teacher survey, administered by study staff.
8. *Summer 2014*. Extant records for all students in participating teachers' classes at the time of state test administration.

Teacher Knowledge Assessment

Reliable measurement of teacher content knowledge is critical to the proposed study and represents the most proximal outcome of the intervention. We will measure teachers' content knowledge at baseline (summer 2013), after completion of the Intel course in the fall 2013, and at the end of the school year in June 2014. We will measure teacher content knowledge with a mathematics assessment composed of items in the Massachusetts Test for Educator Licensure (MTEL) for elementary math teachers. We will draw items from the two MTEL assessments: the mathematics subtest of general elementary test (MTEL#03) and the elementary mathematics assessment (MTEL#53). The MTEL assessments were designed and items validated against a set of test objectives developed and reviewed by practicing educators and faculty at educator preparation institutions. Reported reliability for the MTEL is decision consistency in a licensing context (where the most important outcome is the pass/fail decision). The decision consistency for the general elementary test is 0.92 (on a 0 to 1 scale).

We will create three forms of the teacher knowledge test with 30 items selected from the two MTEL math assessments. Items will be selected that align with topics covered in the Intel Math program including foundations and meanings of addition, subtraction, multiplication, and division; connections between and among addition, subtraction, multiplication, and division, operations with fractions, linear relations, and functions. Selected items will cover specific content (such as items that tap the additive inverse or the meaning of fraction multiplication), and items that require making connections among concepts (such as an item that taps both operations with negative numbers and reducing fractions). Each 30-item form of the teacher knowledge test will require no more than 60 minutes to complete and will be pilot tested prior to administration. See Exhibit 6 for the alignment between the topics covered in Intel Math and the item pool from

MTEL #03 and MTEL #53. There are one or more items from the MTEL assessments that align with each of the 8 Intel units.

Exhibit 6. Alignment of Intel Math Content and the Teacher Knowledge Item Pool (MTEL)

Content Covered in Intel Math Units	Content Covered in Teacher Knowledge Assessments	
	MTEL (#53)	MTEL (#03)
1	5, 12, 42, 44, 50	24, 1
2	1,7	2
3	10, 6, 17	11, 12, 13, 14, 15,17
4	13, 14, 15, 16	
5		18
6	9, 19,18, 20, 21, 22, 23, 24,46	3, 6, 7, 8, 9,10, 23,20, 31, 34
7	45,52, 54,57,62, 63,68, 69,72	28, 29, 30
8	11,43	

In terms of administration, obtaining a measure of teacher knowledge that is completely exogenous to the study and the PD intervention is a critical aspect of the study. The study team will administer the baseline test in person to all teachers prior to random assignment in summer 2013. This will enable the study to avoid a previously-observed phenomenon known as the “late pretest problem” in which scores on a baseline measure taken after random assignment are affected by treatment status (see Schochet, 2008). The baseline teacher knowledge test will include a couple items at the end that ask teachers about their teaching background.

We will administer the first follow-up measure of teacher knowledge in person in fall 2013 (after the completion of the Intel course), and the second in June 2014. For all three rounds of teacher knowledge test data collection we will follow the procedures we have used in other studies (e.g., *The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement*, Garet et al., 2008; *Middle School Mathematics Professional Development Impact Study*, Garet et al., 2011) to train test proctors and monitor the delivery and the secure transmission of study data. The MTEL items are proprietary and therefore the teacher knowledge instruments are not included in Appendix B.

Teacher Survey

The teacher end-of-year survey is included in this package in Appendix B: B-1. The items on this survey are mapped to the constructs that they measure in Exhibit 7. The teacher survey will be administered by study staff at the time of the June 2014 teacher knowledge test.

Exhibit 7. Mapping of End-of-Year Teacher Survey Items to Constructs

Teacher Survey	
Constructs	Items
Professional Development Experiences Related to Math and Math Teaching and Learning	1-8
Math Instruction	9-10
Beliefs About Math Teaching and Learning	11
Certification, Education and Experience	12-15
Demographics	16-18
Perceptions of Study PD Program (Treatment Teachers Only)	19

Extant Data Request

The study team will request administrative records for all students who are in the classrooms of the participating teachers at three time points: (1) summer/fall 2013 when classroom assignments are formed; (2) in March 2014; and (3) at the time that the spring 2014 state assessment is administered. For all students in participating teachers' classes at these three points, the study team will request the following data for both the 2012-13 and 2013-14 school years:

- Demographic characteristics (e.g. gender, race/ethnicity)
- English language learner status, special education status, and free- or reduced-price lunch status
- Math achievement scores on the state assessment

The study team will work closely with district liaisons and districts' data offices to reduce the burden of these requests as much as possible. AIR will facilitate the process of data transfer via secure FTP sites. Exhibit 8 maps the requested extant data fields to constructs; the extant data request protocol is included in Appendix B: B-2.

Exhibit 8. Mapping of Requested Extant Data to Constructs

Extant Data Request	
Constructs	Items
Identifiers and Linkage	1-2A
Student Background Variables	2B
Student Achievement Variables	2C

Classroom Observations

We will measure teachers' classroom practice using the Mathematical Quality of Instruction (MQI) instrument, which is a previously-validated observation protocol applied to video-

recorded observations of teachers' lessons (no burden). AIR will work with a study liaison in each district to hire a part-time person who will be responsible for monitoring consent - procedures and setting up the thereNow HD Insight 2 video cameras in classrooms and securely transferring the data according to a schedule determined collaboratively among an AIR site coordinator, the district liaison, and the participating teachers. Each part-time data collector will be trained via two 2-hour webinars provided by thereNow on how to upload data securely and efficiently. Once uploaded, the data will be transcribed by a pool of Harvard transcribers and passed on to Harvard's deep pool of experienced raters who will code the observation segments according to the MQI.

Study Administered Student Test

Trained study team members will administer the student computer adaptive test to 10 randomly selected students per classroom (stratifying by gender). We will draw the student sample by randomly selecting 10 students per teacher based on the updated records requested by each district in March 2014. The test will include items from Northwest Evaluation Association's (NWEA's) Measures of Academic Progress (MAP). This assessment has an expanding bank of items closely aligned to CCSSM and includes conceptual items that are emphasized in Intel Math. The assessment will focus on a sub-set of topics that are central to Intel Math (e.g., fractions, multiplication and division concepts).

3. Procedures to Maximize Response Rates

Based on our extensive experience with administering surveys in a variety of schools, districts, and states, including a number of similar-size randomized controlled trials, we anticipate a response rate of 85 percent for teacher surveys (relevant randomized controlled trials: *Middle School Mathematics Professional Development Impact Study*, *Early Reading Professional Development Impact Study*, *Online Credit Recovery Study* and *Access to Algebra I Study*). We anticipate response rates above 90 percent on the teacher knowledge assessment and classroom observations due to in-person administration and the negligible burden of the observations, and a 100 percent response rate for the archival records requests.

The studies referenced above had the following response rates for teacher surveys and teacher knowledge assessments (Exhibit 9):

Exhibit 9: Teacher Response Rates from Selected Prior AIR Studies

Study	Teacher Survey Response Rate	Teacher Knowledge Test Response Rate
Middle School Mathematics Professional Development Impact Study	97%	97%
Early Reading Professional Development Impact Study	85%	91%
Online Credit Recovery Study	94%	NA
Access to Algebra I Study	95%	NA

The study team will attend to the following aspects of instrumentation and data collection procedures to ensure high response rates.

- Obtaining high response rates depends in part on the quality of the instruments. The team will pilot and subsequently refine all instruments to ensure that they are user-friendly and easily understandable; this will increase participants' willingness to participate in the data collection activities and thus increase response rates. See the next section for information on piloting procedures designed to ensure instrument quality.
- Obtaining high response rates also depends in part on providing instruments that are a reasonable length. The district screener and teacher survey each require an administration time of approximately 30 minutes. The teacher knowledge assessment requires approximately 60 minutes to complete. These are all reasonable based on similar successful administrations from prior IES studies.
- AIR staff has extensive experience in conducting extant data requests and will facilitate the process by creating a clear and easy to follow request and by following up the initial request with emails and phone calls as necessary. AIR will establish secure FTP sites for districts to upload the requested data securely.
- The study will offer a social incentive to the respondents by stressing the importance of the data collections as part of a high-profile study that will provide much-needed information to the districts and the schools.

4. Pilot-Testing Instruments

The district-level screening protocol was pilot-tested with a small sample of respondents (fewer than 10) for two purposes—to ensure that the instrument and procedures work effectively, and to sharpen estimates of the respondent burden. Based on these considerations, the screener was pilot-tested by telephone with a convenience sample during November 2012. The individuals who participated in the pilot-testing included a district math coordinator and a district curriculum coordinator. Upon completion of the pilot-testing, the items on the screener were found to function as expected, and the time required to complete the screener questions was accurately estimated.

5. Names of Statistical and Methodological Consultants and Data Collectors

This project is being conducted under contract to the U.S. Department of Education by AIR. Michael Garet is a co-Principal Investigator; Jessica Heppen is the Project Director; and Kirk Walters is the Deputy Project Director. Geoffrey Borman from Measured Decisions is a co-Principal Investigator with Garet. Senior task leaders from AIR contributing to the study methods and data collection are Anja Kurki, Toni Smith and Julia Parkinson. For activities associated with the classroom observations, the project includes a subcontract to Harvard University and Clowder Consulting. Key Harvard staff members include Heather Hill and Corinne Herlihy; the key staff member from Clowder Consulting is Catherine McClellan. Project

staff will also draw on the experience and expertise of a network of outside experts who will serve as our technical working group (TWG) members. Prospective TWG members must be approved by IES and are still to be determined.

References

- Garet, M, Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). The impact of two professional development interventions on early reading instruction and achievement (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., et al. (2011). Middle school mathematics professional development impact study: Findings after the second year of implementation (NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance.
- National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education.
- National Science Board. (2012). Science and engineering indicators 2012 (NSB 12-01). Arlington, VA: National Science Foundation.
- Schochet, P. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational Behavioral Statistics*, 33(1), 62–87.
- Spybrook, J., Raudenbush, S. W., Congdon, R., & Martinez, A. (2009). *Optimal design for longitudinal and multilevel research: Documentation for the Optimal Design Software V.2.0*. Available at www.wtgrantfoundation.org