

# i3 Study Design Summary Template

Valid36

Principals on the Path to Excellence  
(National Institute for School Leadership)

## Contents

{ TOC \o "1-3" \h \z \u }

## 1. Introduction

### 1.1. Evaluator

Dr. Steven M. Ross, Ph.D.  
Professor  
School of Education  
Johns Hopkins University

300 East Joppa Road—Suite 500  
Baltimore, MD 21286

Phone: 901-573-2479

{ HYPERLINK "mailto:roisin.corcoran@aya.yale.edu" }

### 1.2. Intervention background/history

The intervention for this study is the National Institute for School Leadership's (NISL) Executive Development Program (EDP) with on-going coaching. This program is a comprehensive professional development initiative designed to improve leadership knowledge and skills of school principals across the United States. Through classroom instruction, simulations, interactive online sharing of ideas, and other resources, the EDP provides for a yearlong, cohort-based professional development program that includes three courses and 12 units.

The EDP is the culmination of a four-year, \$11 million R&D initiative undertaken by the National Center on Education and the Economy (NCEE), NISL's parent company, with strong philanthropic support from Carnegie Corporation of New York, The Broad Foundation, the New Schools Venture Fund and the Stupski Foundation. The R&D effort included benchmarking the best educational leadership development practices worldwide and identifying the best adult learning methods and strategies used in business, medicine, law, education, and the military.

Exceptional leadership development approaches culled from international best practices, leading professions, and the military now permeate the EDP. The program offers sustained, cohort-based, job-embedded, applied learning, using a blended learning model of face-to-face and technology-enhanced learning, including video interviews from leading educators, case studies, best practice videos, game playing, simulations, and journaling.

In 2005, after a successful pilot of the EDP, NCEE launched the National Institute for School Leadership (NISL) to manage the implementation and scale-up of the program. NISL's proven train-the-trainer model, in which state and district leaders are fully trained in the EDP and certified as facilitators to deliver the program with fidelity, builds state and district capacity to take ownership of the training at a cost that is affordable and sustainable. More than 13,000 school leaders in 27 states have successfully completed the program.

The EDP is closely aligned with the most current and highly regarded standards for high-performing principals, including the Interstate School Leaders Licensure Consortium (ISLLC) Standards (Council of Chief State School Officers, 2008) and Public Impact's competencies for turnaround leaders (Public Impact, 2008). NISL updated the EDP in 2010 and again in 2013 to incorporate the latest research on leadership and to align with the underlying concepts that were built in to the Common Core State Standards (e.g., the cognitive science research on how people learn and the empirical evidence of what knowledge and skills are required to be ready for college and work).

However, the key building blocks of the EDP—educational best practices, instructional leadership skills, subject-area knowledge, and dynamic adult pedagogy—have remained intact.

In terms of prior empirical investigations into the overall program’s efficacy, several key research projects have been conducted thus far. First, NISL has been running a successful state-wide novice principal program for the past six years in Pennsylvania and has already demonstrated its positive impact on student achievement. Since 2008, Pennsylvania law has required all novice principals to complete a rigorous leadership development program. The only program that qualifies to meet this requirement for state funding is the EDP, the foundation of this proposal. Independent researchers used a rigorous methodology to evaluate the Pennsylvania EDP implementation and found statistically significant gains in student learning in English language arts (ELA) and mathematics as measured by state test scores (Nunnery, Yen, & Ross, 2011). The effect sizes were .08 and .07 standard deviations in ELA and mathematics, respectively. This finding translates to roughly one to two months of additional learning on average for the 57,000 students in 101 Pennsylvania treatment schools. Another way to measure student learning is the percentage of students who reach the state proficiency performance levels. In the Pennsylvania study, researchers found that 2.16% more students achieved ELA proficiency in treatment schools than otherwise would be expected and 1.92% more did so in mathematics. This finding translated to 1,225 more proficient students in ELA and 1,089 more in mathematics.

NISL also has run a statewide leadership development program in Massachusetts since 2006. Whereas the Pennsylvania project focuses on novice principals across all schools, the Massachusetts project focused on principals in high-need schools, including novice principals. Using a rigorous methodology, researchers from Old Dominion University and Johns Hopkins University evaluated the results of the second round of training (Nunnery, Ross, Chappell Moots, Pribesh, & Hoag-Carhart, 2011). The researchers found a statistically significant impact on student achievement in both ELA and mathematics. The effect sizes (in standard deviations) were .11 in ELA

and .14 in mathematics for the 21,000 students in 38 Massachusetts treatment schools (average poverty level of 69%).

Finally, one study identified seven exemplar leadership training programs (Cheney, Davis, Garrett, & Holleran 2010); all costing between \$100,000 to \$200,000 per graduate. The EDP has produced better student achievement results than all seven programs for an average of \$10,000 to \$25,000 per graduate (depending on implementation design). This result makes this project particularly important: It will validate a different approach to school leadership that produces stronger results for a fraction of the investment.

We have found a strong correlation between the number of key concepts that a principal implements after completing the EDP and student achievement gains. In fact, student achievement gains doubled for principals who were identified as more aggressive implementers versus the average incremental gain (The Meristem Group, 2009). Therefore, our project includes a highly focused coaching model designed to increase the implementation of key EDP concepts. All 135 novice principals in the treatment schools will be provided with this additional support during the program. NISL faculty will provide this coaching, using the best practices in the EDP as a focal point for school change. This coaching support—a total of 11 days per principal—will be delivered over 15 months in a combination of in-person, Skype, telephone, and email.

### **1.3. Confidentiality protection**

The outlined study, *Principals on the Path to Excellence: A Validation Study of the National Institute for School Leadership Executive Development Program*, has been reviewed and approved by the Johns Hopkins University Homewood Institutional Review Board (Protocol/FSA Number: HIRB00002940). As such, the procedures and methods utilized by this project meet the federal regulatory standards of the US Department of Health and Human Services Office for Human Research Protections (OHRP).

Further district-level research approvals mandated in certain participating districts have also been submitted and are pending approval. The following IRB-approved procedures will be utilized to protect the confidentiality of data.

Achievement and demographic data collected from the states or districts as well as focus group comments, interviews, surveys, and other materials collected from participants will be maintained by researchers in coded form, with names and other identifiers replaced by a code number, stored electronically and password protected. Participant names will be maintained in a separate codebook, with access limited to research personnel and password protected.

Access to the electronic folder where all data is stored is limited to research personnel via their individual log-in and password in addition to the password protection applied to individual electronic folders. All computers used by research personnel are subject to security safeguards including password-protected computers, antivirus controls, malware protection, firewall configuration, and scheduled automatic backups. Hard copies of all de-identified data will be kept in a locked file cabinet in a secured office. Based on Homewood IRB research protocol, all data will be kept for three years following the completion of the study or until all research participants reach adulthood (18 years of age).

Finally, the research team made up of researchers from Johns Hopkins University (JHU), Old Dominion University, and NISL will be responsible for the collection and storage of data. Only this research team will have access to the research data.

#### ***1.4. Independence of evaluation***

Johns Hopkins University (JHU) will be the independent evaluator of the NISL EDP program and coaching model. In this role, JHU will independently conduct all key aspects of the evaluation. Specifically, JHU will be responsible for collecting and analyzing data from both the treatment and

control conditions in order to estimate program impacts on outcomes for students, principals, and teachers. JHU will gather data needed to (a) provide ongoing performance feedback to NISL and the Project Coordinating Committee, (b) execute a randomized experimental research design to determine the impact of the EDP, (c) evaluate fidelity of implementation, and (d) develop case studies of the EDP program in each of three state sites. Findings reported to NEi3 will be subject to review but not approval of the project developer/grantee.

## 2. Brief Summary of Intervention and Evaluation Design

Type of grant: Validation grant

The intervention consists of NISL's Executive Development Program for School Leaders (EDP) augmented with sustained one-on-one coaching focused on deepening the lessons of the EDP. Each 20–30 principal cohort will be assigned two NISL Master Faculty members, who have previously received extensive training and have experience delivering the EDP. A "principal cohort" is a group of principals who are joined together as an EDP class throughout the delivery of all 12 units, generally over a span of 12 or more months. Each cohort (there will be three in each of the three states) will consist of the treatment principals located within approximately two hours driving distance of the meeting site. Members of each cohort will meet approximately once per month, for two days at a time, for approximately one year to complete the 12 units of the EDP. In between the monthly in-person sessions, participants of a cohort will have the opportunity to interact with each other via NISL's online Learning Management System (LMS). Twenty-four full days of training will be conducted over 12-14 months in two-day increments that correspond to each unit of the EDP. In addition to this in-person work, EDP participants will be required to complete several hours of prework and reading to prepare for each session.

During the EDP, each participant, with guidance from the facilitators and coaches, will design an Action Learning Project (ALP) built around a high-priority issue at his or her school. As the EDP proceeds, the ALPs are designed to both yield real benefits to the participating schools, but also offer other cohort members a set of real-world case studies to learn from. Following Unit 3, as the ALPs are being finalized, each participant will be matched with a NISL-trained coach. NISL's coaching program will provide 11 total days of one-on-one coaching, focused on the EDP content and the implementation of participants' ALPs. Nine of these 11 days of coaching will be delivered in person, with those hours mostly delivered as half-day, in-person visits. The remainder of the coaching hours

will be fulfilled via email, phone, and Skype communication. All 11 days are expected to be delivered within a period of 18 months, with 3 falling in SY 2015-16 and 8 in SY 2016-17. The EDP begins in fall of the 2015-16 academic year and continues through fall of the SY 2016-17. The coaching begins during the EDP (after Unit 3) and will continue through SY 2016-17. In summary, coaching time will include half-day in-person visits to the participants' schools, in addition to time spent coaching via phone, video conference, and email communications. Coaches and coachees will communicate at least once per week for the approximately 15 months of coaching, which will extend to the close of the 2016–2017 school year.

*Description of all impact studies being conducted*

The independent evaluation of the impact of the NISL EDP will be led by Dr. Steven Ross at the Center for Research and Reform in Education (CRRE) at Johns Hopkins University (JHU). The purpose of the proposed study is to investigate the impact of the EDP and coaching on a variety of student, school, and teacher outcomes. To do so, the independent evaluators will conduct a multisite cluster-randomized trial across three states (California, Florida, and Mississippi). Within each state site, schools/principals were randomly assigned to either the treatment (EDP and coaching) condition or to the control (business-as-usual) condition.

*The key research question(s) for the impact study.*

The main confirmatory analysis will estimate the impacts of the EDP and coaching on student achievement in mathematics and in reading across all three state sites and after three years of program implementation. Additional exploratory analyses will estimate separate impacts of the EDP on student achievement for each state site. For all analyses of the impacts of the EDP on student outcomes, students are clustered within schools.

The main confirmatory analyses use an intent-to-treat approach in which schools remain in the study regardless of whether the principal remains in the school or whether principals who were assigned to

the treatment condition participate in the EDP over the three-year period. The evaluators will also explore the impacts of the EDP in schools where principals were retained over the three-year period and for principals who participated in the EDP if they were randomly assigned to the treatment condition; these additional analyses will be exploratory. Finally, additional exploratory analyses will estimate the impact of the EDP on students' science achievement, school conditions, and teacher retention.<sup>1</sup>

The impact study will seek to address the following research questions:

**Confirmatory Research Question**

1. What is the impact of EDP for novice, middle school principals on students' mathematics and reading achievement compared with the business-as-usual condition for novice, middle school principals after three years of program implementation?

**Exploratory Research Questions**

2. In each of the three state sites, what is the impact of EDP for novice, middle school principals on students' mathematics and reading achievement compared with the business-as-usual condition for novice, middle school principals after three years of program implementation?
3. What is the impact of three years of EDP participation (rather than EDP access) for novice, middle school principals on students' mathematics, reading, and science achievement compared with the business-as-usual condition for novice, middle school principals?
4. What is the impact of two years of EDP participation for novice, middle school principals on students' mathematics and reading achievement compared with the business-as-usual condition for novice, middle school principals?

---

<sup>1</sup> The analysis of teacher retention is subject to the availability of the necessary teacher human resource data collected from state administrative databases.

5. What is the impact of three years of EDP participation for novice, middle school principals on students' mathematics and reading achievement for various student subgroups (e.g., students living in poverty, English language learner (ELL) students, and students with disabilities)?
6. Does the impact of three years of EDP participation for novice, middle school principals on students' mathematics and reading achievement vary for higher and lower achieving schools? What might these differential effects be attributable to?<sup>2</sup>
7. What is the impact of three years of EDP participation for novice, middle school principals on school conditions, such as school leadership, teacher collaboration, family involvement, supportive environment, and ambitious instruction compared with the business-as-usual condition for novice, middle school principals?
8. What is the impact of three years of EDP participation for novice, middle school principals on the school-level teacher retention rate compared with the business-as-usual condition for novice, middle school principals?<sup>3</sup>

#### The study design

The impact portion of the evaluation will involve a multisite cluster-randomized control trial in three states (California, Florida, and Mississippi). Within each state site, schools/principals were randomly assigned to either the treatment (EDP and coaching) condition or to the control (business-as-usual) condition. The main confirmatory analyses will estimate the impacts of the EDP and coaching on student achievement in mathematics and in reading across all three state sites and after three years of program implementation.

#### The study sample

For the purposes of the i3 program, NISL has chosen to target middle school principals with five or fewer years of experience in the principalship. NISL's recruitment strategy targeted districts with 50% or greater free and reduced-price lunch eligibility in order to recruit a sample of districts with a high

---

<sup>2</sup> The latter part of this research question will be addressed using qualitative data.

proportion of low-income students. As a result, the vast majority of the principals in the i3 study will be working in schools with greater than 50% free and reduced-price meal-eligible students.

Over 100 school districts across the three states were selected for participation. These districts were selected because they are interested in and committed to carrying out the intervention and evaluation/research component, serve student populations that are diverse in terms of ethnic/racial background and socioeconomic status, have a demonstrated need to improve academic achievement, and allow for principal cohorts to be formed without requiring excessive travel time for treatment principals to attend the in-person trainings.

From each of the three states to be included in this study—California, Florida, and Mississippi—we selected approximately 100 middle schools (or schools of other configurations including grades six through eight) to participate in the study. Within each state and, where possible, district, schools were randomly assigned. Prior to random assignment of schools, we created matched pairs of schools within each of the three states based on district membership, the percentage of students achieving proficiency in mathematics and reading over multiple years, and principal years of experience. Within each matched pair of schools, schools were randomized—one to the EDP and coaching treatment and one to the control condition. We will not, however, use the matched school pairs in the impact analysis; instead, we will analyze outcomes for all participating schools and control for state, district, and school-level characteristics used to create matched pairs.

Finalization of school/principal selection occurred in September 2015. The final sample of schools is as follows: 80 schools in Mississippi (40 treatment and 40 control), 86 schools in Florida (43 treatment

---

<sup>3</sup> This exploratory research question is subject to the availability of the necessary teacher human resource data collected from state administrative databases.

and 43 control), and 104 schools in California (52 treatment and 52 control) for a total of 270 schools (135 treatment and 135 control).

For the confirmatory analyses, the impacts of the EDP on students' mathematics and reading achievement will be measured after three years of program implementation for one cohort of students (for students who were in the sixth grade in study schools before random assignment of schools).

Achievement outcomes will be measured for these students at the end of their eighth grade year (spring 2018), after three years in a study school.

*The expected outcomes*

Based on the results of previous research conducted on the NISL EDP program, the EDP program is expected to improve both student and school outcomes. This study will examine impacts of the EDP program on students' mathematics, reading, and science achievement; school conditions, such as school leadership, teacher collaboration, family involvement, supportive environment, and ambitious instruction; and teacher retention.

*The analysis approach(es) that will be used to answer research questions (e.g., multilevel modeling, OLS regression).*

The confirmatory and exploratory questions regarding the impacts of EDP on student achievement will be addressed using hierarchical linear modeling (HLM) where students are nested within schools. The treatment effect will be estimated at the school level, reflecting the assignment of schools to either the treatment or control condition.

*Description of implementation study, including:*

*A brief narrative of the theory of change for the intervention (based on the logic model).*

Exhibit 4.1. NISL Executive Development Program – i3 Logic Model (see page 38) outlines the theory of change proposed by the NISL EDP program. As demonstrated in this figure, the combination of NISL EDP curriculum and coaching leads to the refinement of principal leadership practices, which in

turn leads to changes with school policy, culture, and practices within the principal's school. Finally, as a prospective result of these school-based changes, student outcomes are influenced.

*The research question(s) for the implementation study.*

The implementation study will involve collecting data in order to monitor and analyze the delivery and perceived quality of implementation of the EDP and coaching in each of the three states. The specific research questions which will be addressed include:

**Fidelity of Implementation (FOI) Questions**

1. Across all three states, is each of the key components of the EDP intervention implemented with fidelity for the full sample of schools?
2. For each state site, what percent of NISL schools have a high FOI according to the FOI matrix?

**Program Quality Questions**

3. How do principals and coaches perceive the effectiveness of the EDP? How has the EDP changed principals' priorities, practice, and how they allocate their time?
4. How do principals perceive the quality of the support they received through coaching?

*Plan for measuring fidelity of implementation (FOI).*

Findings about implementation fidelity, as scored using the FOI matrix (pages 48-49), will be reported for each of the two school years that the EDP intervention is implemented.<sup>4</sup> The FOI will be reported separately for each program component across all three states and for each state separately. The FOI will include an indication of whether or not the implementation of each program component met the established threshold for "implementation with fidelity" at the sample level. The implementation study section towards the end of this document provides additional details.

---

<sup>4</sup> The EDP training and coaching occur for only two years. The main confirmatory contrast focuses on student achievement outcomes in the third year of the study when fully-trained principals are leading schools.

Data collection instruments.

Data will be collected from a variety of sources to assess the FOI, or the extent of fidelity of program delivery and principal participation in the program. Data collection areas that will be targeted include: (a) principal attendance to EDP group trainings captured by attendance logs, (b) the number of coaching days delivered captured by principal and coach logs, and (c) whether principals received all of the EDP materials captured by shipping company records of document delivery of training materials. Data will also be collected to gauge principal and coach perceptions of EPD program quality and effectiveness. Data collection efforts include the following.

*Training surveys.* Each training session will be evaluated by NISL using a brief survey asking participants to rate clarity, organization, and relevance in relation to the logic model's direct outcome dimensions. These surveys, which include Likert-type ratings items and open-ended questions, are used routinely by NISL in its EDP trainings. The surveys will be administered online by NISL.

*Training observations.* JHU will draw a purposeful sample of nine training sessions (three per state) to be observed. In each site, JHU will record impressions using an informal "observation guide" designed to capture (through open and prompted field notes) key dimensions of the training (e.g., organization, clarity, and time management), adaptation to diversity (school characteristics, principal experiences), relevance to session goals, and participant engagement. This observation guide will be developed by CRRE in collaboration with NISL.

*Principal focus groups.* At the completion of the EDP training and again at the completion of coaching, six focus groups (two per state, each consisting of six-to-eight randomly selected principals) will be conducted by webinar to determine reactions to the EDP implementation (i.e., curriculum content, professional development activities, and coaching) in relation to the logic model outcomes.

*Coach focus groups.* Three focus groups (one per state) of randomly selected coaches will be conducted by conference call at the completion of coaching. Questions will focus on principals'

participation, their interest in such guidance, the ability to capitalize on it, their readiness to take risks, and their ability to capitalize on what the EDP offers them.

*NISL principal survey.* At the end of each school year, principal participants will be asked to complete a survey regarding their experiences in the EDP program and its applications to leadership activities in their schools (for treatment principals only). Treatment and control principals will also be surveyed about the actions they have taken to improve the quality of education students are receiving and to improve the working environment and culture for teachers and students. Survey data will also be used to determine and monitor the diffusion of treatment to principals in the control condition.

*Principal shadowing.* For the purposes of the study, an “EDP Principal Shadowing Tool” will be developed by CRRE in consultation with NISL for use in six case study schools. The tool will be designed to collect reliable and rich qualitative data on principals’ everyday practices relative to and separate from their EDP training. It will involve a guided conversation and reflections between the researcher and the principal as the latter goes about a typical school day. Researchers will also interview principals at the six case study schools.

*Analysis approach.* The implementation study will make use of data collection efforts described above. In addition, JHU will conduct case studies in six schools (two per state) to shed light on how EDP novice principals use their new knowledge and skills to change key aspects of how their schools function to increase student achievement. These schools will be selected from the population of treatment schools. For each case study school, JHU will visit the school three times over the course of the study (Fall 2016, Fall 2017, and Fall 2018), for a total of 18 visits for all six case study schools. In addition to the data collection measures used for all treatment and control schools, additional data on principals in case study schools will be collected by the research team using the “Principal Shadowing Tool” and interviews.

The analyses of qualitative data will be guided by Miles and Huberman's (2004) model, consisting of transcribing the responses, deriving codes, identifying themes, and revision and refinement based on member checking and inter-rater review. Triangulation across data sources and methods will be used to validate the major findings. Cross-case analyses, member-checked through an online focus group of key informants (Miles & Huberman 2004), will be used to identify programmatic, individual, and contextual factors. Each case study will be member-checked with key informants. Coding for quality assurance will be applied to the analyses of qualitative data.

Specifically, CRRE analysts will subject the qualitative data sources to coding using NVivo, a software program that facilitates coding, analysis, and reanalysis of qualitative data in multiple file formats.

NVivo is a system created for mixed-methods research that allows multiple users to collaborate by viewing and accessing qualitative and quantitative data within an Internet-based system. NVivo enables researchers to upload files for coding, store and manage data securely, link qualitative and quantitative data, and identify data patterns for further analyses. Ultimately, using these tools and techniques, two principles will be utilized to analyze these data: (a) triangulation and (b) grounded theory. Coding the data in this way typically leads to a number of outcomes (for further discussion, see Bryman, 2008, p. 544), which are as follows: concepts, category, hypotheses (initial hunches about relationships between concepts); and theory (a set of well-developed categories that form a theoretical framework).

### 3. Impact Evaluation

List each of the impact studies in the table below.

Impact studies described in Chapter 3		
Chapter Section	Title	Notes
3.1	Impact Study	<i>Impact of the NISL EDP intervention on student, teacher, and school outcomes using a multisite cluster-randomized trial</i>

### 3.

#### 3.1. Impact Study

##### 3.1.1. Introduction to Impact Study

Provide a very brief overview of the design of the impact study here.

- In particular, is the study design an RCT, a QED, an RDD, or a prepost design (for development grants only)?

This study is a multisite cluster-randomized control trial in three states (California, Florida, and Mississippi) in which schools were randomly assigned to either the treatment condition or to the business-as-usual (control) condition.

*How are the treatment and comparison groups defined?*

For the purposes of the i3 program, NISL has chosen to target middle school principals with five or fewer years of experience in the principalship. NISL's recruitment strategy targeted districts with 50% or greater free and reduced-price lunch eligibility in order to recruit a sample of districts with a high proportion of low-income students. As a result, the vast majority of the principals in the i3 study will be working in schools with greater than 50% free and reduced-price meal-eligible students.

Over 100 school districts across the three states were selected for participation in three state sites (California, Florida, and Mississippi). These districts were selected because they are interested in and committed to carrying out the intervention and evaluation/research component, serve student

populations that are diverse in terms of ethnic/racial background and socioeconomic status, have a demonstrated need to improve academic achievement, and allow for principal cohorts to be formed without requiring excessive travel time for treatment principals to attend the in-person trainings. Within each state, districts were recruited within geographic regions that would allow for cohorts to be formed without requiring excessive travel time for treatment principals to attend the in-person trainings.

In interested districts, all schools containing grades six through eight that were led by principals with five or fewer years of experience in the principalship were considered eligible for the study pool.

Extremely small schools, generally those with fewer than 10 students per grade, were excluded as they generally indicated unique school structures (i.e. virtual schools, etc.). Principals who had prior exposure to NISL's EDP were also not included in the study.

Schools were matched in pairs based on district membership (where possible), the percentage of students achieving proficiency in mathematics and reading over multiple years, and principal years of experience; then, schools within each matched pair were randomly assigned to either the treatment or control condition. Half of the principals in participating schools will be given access to the EDP and coaching model (treatment), and half will receive the business-as-usual district professional development for novice principals (control). We will not, however, use the matched school pairs in the impact analyses; instead, we will analyze outcomes for all participating schools and control for state, district, and school-level characteristics used to create the matched pairs.

Finalization of school/principal selection occurred in September 2015. The final sample of schools is as follows: 80 schools in Mississippi (40 treatment and 40 control), 86 schools in Florida (43 treatment and 43 control), and 104 schools in California (52 treatment and 52 control) for a total of 270 schools (135 treatment and 135 control).

- *What are the primary outcomes being examined?*

Outcome variables to be examined include:

1. Student achievement on state standardized assessments in mathematics, reading, and science.
2. School conditions such as school leadership, teacher collaboration, family involvement, supportive environment, and ambitious instruction.
3. School-level teacher retention rate.<sup>5</sup>
4. Cost-effectiveness of the EDP program in terms of impacts of the EDP program on student achievement relative to instructional expenditures.<sup>6</sup>

*State the filename of the Contrast Table that contains information on the research questions, outcomes, baseline measures, sample, and contrasts relevant to the study being described in this section.*

Valid36\_Contrast\_Tool\_5-2-16.xlsm

### **3.1.2. Research questions**

#### **Confirmatory Research Question**

1. What is the impact of EDP for novice, middle school principals on students' mathematics and reading achievement compared with the business-as-usual condition for novice, middle school principals after three years of program implementation?

#### **Exploratory Research Questions**

2. In each of the three state sites, what is the impact of EDP for novice, middle school principals on students' mathematics and reading achievement compared with the business-as-usual condition for novice, middle school principals after three years of program implementation?

---

<sup>5</sup> This exploratory outcome is subject to the availability of the necessary teacher human resource data collected from state administrative databases.

<sup>6</sup> NISL is responsible for producing the cost-effectiveness analysis.

3. What is the impact of three years of EDP participation (rather than EDP access) for novice, middle school principals on students' mathematics, reading, and science achievement compared with the business-as-usual condition for novice, middle school principals?
4. What is the impact of two years of EDP participation (rather than EDP access) for novice, middle school principals on students' mathematics and reading achievement compared with the business-as-usual condition for novice, middle school principals?
5. What is the impact of three years of EDP participation for novice, middle school principals on students' mathematics and reading achievement for various student subgroups (e.g., students living in poverty, English language learner (ELL) students, and students with disabilities)?
6. Does the impact of three years of EDP participation for novice, middle school principals on students' mathematics and reading achievement vary for higher and lower achieving schools? What might these differential effects be attributable to?<sup>7</sup>
7. What is the impact of three years of EDP participation for novice, middle school principals on school conditions, such as school leadership, teacher collaboration, family involvement, supportive environment, and ambitious instruction compared with the business-as-usual condition for novice, middle school principals?
8. What is the impact of three years of EDP participation for novice, middle school principals on the school-level teacher retention rate compared with the business-as-usual condition for novice, middle school principals?<sup>8</sup>

### **3.1.3. Control or comparison condition**

*Provide a description of the control or comparison condition, as well as any confounds that could bias the estimated effect of the intervention.*

---

<sup>7</sup> The latter part of this research question will be addressed using qualitative data.

<sup>8</sup> This exploratory research question is subject to the availability of the necessary teacher human resource data collected from state administrative databases.

Control schools/principals will not receive any NISL services or exposure to EDP materials prior to completion of the study. Principals in control schools will have access to the business-as-usual supports for novice principals available in their districts. We will minimize the threat to internal validity posed by potential diffusion of treatment through the following mechanisms:

- NISL has fully explained the ramifications of engaging in a randomized controlled trial to the Official Partner districts and will secure commitments from additional participating districts that they will adhere to protocols to prevent diffusion of treatment.
- Treatment principals will be asked to sign a letter stating that they understand that dissemination of EDP materials would be a violation of the terms of the district contract agreement with NISL.
- An annual survey of control school principals will be administered to monitor possible diffusion of treatment. If potential problems are indicated on the survey, we will conduct phone interviews to determine and document the nature and extent of diffusion.
- Eligible principals with prior access to NISL's EDP were not included in the sample.

With these facets in mind, we do not anticipate confounds to this study.

### **3.1.4. Sample identification, selection and assignment**

*For all designs, in Section 3.1.4, provide a brief overview of the sampling design and process of assignment to treatment and control/comparison conditions.*

For the purposes of the i3 program, NISL has chosen to target middle school principals with five or fewer years of experience in the principalship. NISL's recruitment strategy targeted districts with 50% or greater free and reduced-price lunch eligibility in order to recruit a sample of districts with a high proportion of high-need students. As a result, the vast majority of the principals in the i3 study will be from schools with greater than 50% free and reduced-price meal-eligible students.

Within each state, we pair-matched schools, based on district membership (where possible), the percentage of students achieving proficiency in mathematics and reading over multiple years, and principal years of experience. After matching was completed, participating schools in each matched pair were randomly assigned—one to the treatment group and one to the control group.

#### **3.1.4.1. Identification/selection of study districts**

The study will take place in three states—California, Florida, and Mississippi. In each state, we have recruited 30 to 60 districts, depending on district size, in order to recruit approximately 100 middle schools in each state with novice principals.

Over 100 school districts across the three states have been selected for participation. These districts have been selected because they are interested in and committed to carrying out the intervention and evaluation/research component, serve student populations that are diverse in terms of ethnic/racial background and socioeconomic status, have a demonstrated need to improve academic achievement, and allow for principal cohorts to be formed without requiring excessive travel time for treatment principals to attend the in-person trainings.

#### **3.1.4.2. Identification/selection of study schools**

Within interested districts, all schools containing grades six through eight that were led by principals with five or fewer years of experience in the principalship were considered eligible for the study pool.

Extremely small schools, generally those with fewer than 10 students per grade, were excluded because they generally indicated unique school structures (i.e. virtual schools, etc.). Principals who had prior exposure to NISL’s EDP were also not included in the study.

Prior to random assignment of schools, we created matched pairs of schools within each of the three states, based on district membership (where possible), the percentage of students achieving proficiency in mathematics and reading over multiple years, and principal years of experience. Within each matched pair of schools, schools were randomized—one to the EDP and coaching treatment and one to the business-as-usual control condition. We will not, however, use the matched school pairs in the impact analyses; instead, we will analyze outcomes for all participating schools and control for state, for district, and for the school-level characteristics used to create matched pairs.

Finalization of treatment and control site matching and assignment occurred in September 2015 for Mississippi, California, and Florida sites. The final sample of schools is as follows: 80 schools in Mississippi (40 treatment and 40 control), 86 schools in Florida (43 treatment and 43 control), and 104 schools in California (52 treatment and 52 control) for a total of 270 schools (135 treatment and 135 control).

### **3.1.4.3. Identification/selection of study teachers**

For the analyses of EDP impacts on student achievement outcomes, there is no selection of a teacher sample. For the confirmatory analyses of the EDP impacts on student achievement, the student sample will be comprised of all students in grade six in school year (SY) 2015–2016 who were present in any study school at the time of random assignment.

For the exploratory analyses of the EDP impacts on the school-level teacher retention rate, the outcomes sample will include all teachers in any study school in the 2015-16, 2016-17, or 2017-18 SYs. The baseline sample will include all teachers in any study school in the 2014-15 SY (prior to the EDP implementation). Therefore, the retention analysis will include teacher joiners.

For the exploratory analyses of EDP impacts on school conditions, teachers of grades six through eight will be invited to complete the Consortium on Chicago School Research (CCSR) “Our Schools, Our Voices Teacher Survey.” This teacher survey includes questions regarding school leadership, teacher collaboration, family involvement, supportive environment, and ambitious instruction. This survey will be administered each spring (Spring 2016, Spring 2017, Spring 2018) to all teachers and students in grades six through eight in all study schools; therefore, the teacher survey samples may include those who join the school after random assignment. Data from the teacher survey will provide school-level measures of how well the school functions from the perspectives of teachers.

The survey will also be administered in Fall 2015 to teachers in a randomly selected subsample of schools (approximately 15 treatment and 15 control in each state) in order to generate baseline data (for a subsample of schools) and to field test survey administration methods.

#### **3.1.4.4. Identification/selection of study classes**

All classes in participating schools and in relevant grades will be included in the proposed analyses. As described in the previous section regarding the teacher sample, there will be no selection of a sample of classes within study schools.

#### **3.1.4.5. Identification/selection of students**

##### **Student samples for analyses of EDP impacts on student achievement outcomes**

Two cohorts of students will be followed over the course of the evaluation. The confirmatory analyses will focus on one cohort of students who were in the sixth grade in the 2015–2016 SY, and the study will follow these students for three SYs (2015-16, 2016-17, 2017-18) through their eighth grade year. We will assess impacts of the EDP program on student achievement outcomes in Spring 2018 after three years of program implementation.

Baseline mathematics and reading achievement for Cohort 1 sixth-grade students will be collected from Spring 2015, prior to the start of the intervention and when students were in the fifth grade.

Mathematics and reading assessments will be obtained in each subsequent year with Cohort 1's eighth-grade assessments from Spring 2018 used as the outcomes in the confirmatory analyses after three years of program implementation.

We will also conduct exploratory analyses using a second cohort of students who will be in the sixth grade in the 2016–2017 SY. The study will follow this second cohort of students for two SYs (2016-

17, 2017-18) through their seventh grade year, again assessing outcomes in Spring 2018 but after two years of program implementation.

The table below illustrates the progression of both cohorts of students over time, including the timeline for pretest, intermediate, and posttest data collection. The confirmatory impact analyses will focus on the Cohort 1 sample of students who were in grade six in fall 2015 and who were known to be present in study schools prior to school random assignment. Student rosters were obtained from study schools to identify students present in schools in Fall 2015 prior to school randomization. Thus, there are no student joiners in Cohort 1. Cohort 2 contains student joiners, but analyses with students in Cohort 2 are exploratory.

Table 1: Student Cohorts for the Student Achievement Impact Analyses

	Cohort	2014–15 (Pretest)	2015–16	2016–17	2017–18 (Posttest)
Grade levels Tested	1	5	6	7	8
	2	4		6	7

Cohort 2 will be comprised of a sample of students who were in the sixth grade in Fall 2016, including student joiners. Many of the study schools begin with the sixth grade, thereby making incoming sixth-grade students joiners. However, individual student pretests will be obtained from Spring 2015, prior to school randomization and when students were in the fourth grade. Mathematics and reading achievement scores will be obtained from Spring 2017 and Spring 2018 with Cohort 2’s seventh-grade assessments from Spring 2018 used as the outcomes in the exploratory analyses after at least two years in a study school.

The main confirmatory impact analyses (and most exploratory analyses) will evaluate the effectiveness of the EDP after three full years of implementation. We expect the impacts of the EDP on

student achievement to be strongest after three full years of implementation and when treatment principals complete the EDP training and coaching and remain in their schools over the course of the study.

Schools will average at least two sixth-grade classrooms, yielding from 160 to 208 classrooms per state (208 in California, 172 in Florida, and 160 in Mississippi), with approximately 25 students per classroom, for an estimated 4,000 to 5,200 students per state (2,000 to 2,600 per treatment condition in each state).

### **Student sample for analysis of impacts of EDP on school conditions**

For the analysis of the impacts of EDP on broader school-level outcomes (e.g., school leadership) as reported by teachers and students, all students in grades six through eight in all study schools will be invited to complete the Consortium on Chicago School Research (CCSR) “Our Schools, Our Voices Student Survey.” The student survey includes questions regarding supportive environment and ambitious instruction. This survey will be administered each spring (Spring 2016, Spring 2017, Spring 2018) to all teachers and students in grades six through eight in all study schools; therefore, the student survey samples may include those who join the school after random assignment.

The survey will also be administered in Fall 2015 to students in a randomly selected subsample of schools (approximately 15 treatment and 15 control in each state) in order to generate baseline data (for a subsample of schools) and to field test survey administration methods. Data from the student survey will provide school-level measures of how well the school functions from the perspectives of students.

### ***Inference space (focal population) - optional***

The EDP has demonstrated results in improving student achievement in elementary, middle, and high schools; struggling schools and high performing schools; urban, rural, and suburban schools; and in schools located in different geographic regions. This study evaluates the impact of the EDP program and coaching for novice principals working in middle schools serving predominantly disadvantaged student populations. For this study, we have recruited roughly 100 schools in each of three states to participate. Therefore, this study evaluates the impact and implementation of the EDP across a rich array of locations and settings. In turn, results generated through this study will look to generalize to similar populations of similar scope.

#### ***3.1.4.6. Multiyear interventions***

*If the impact study includes multiple years, describe the way the sampling design will follow schools/teachers/students over multiple years.*

Novice principals in treatment schools will receive the initial intervention training (the NISL EDP) in Fall 2015 continuing through Fall 2016, and coaching will begin in Spring 2016 and will continue through Summer 2017.

As previously discussed, the student sample for the confirmatory analyses will focus on a single cohort of sixth-grade students present in study schools before school random assignment (Cohort 1). Baseline mathematics and reading achievement for these students will be collected from Spring 2015, prior to the start of the intervention and when students were in the fifth grade. Mathematics and reading assessments will be obtained in each subsequent year and students' eighth-grade assessments from Spring 2018 will serve used as the outcomes in the confirmatory analyses. For the confirmatory analyses, student achievement outcomes will be analyzed after three years of program implementation. There are no student joiners for the confirmatory analyses.

### 3.1.5. Attrition (RCTs only)

The evaluators will follow the guidelines proposed by What Works Clearinghouse (WWC) (WWC, 2014) to determine the overall and differential attrition at both the cluster (school) and student levels.

We will use an intent-to-treat model, following all students in all schools randomly assigned at the outset, regardless of whether principals participating in the study remained in their schools. Thus, we expect cluster (or school) attrition to be minimal and within an acceptable level of bias (WWC, 2014). Specifically, over the three-year period, we would expect less than 10% overall school attrition and less than a 5% difference in attrition between treatment and control schools. In addition, because the intervention will be administered to principals in the treatment group, we expect attrition of students to be predominantly exogenous (primarily family mobility) and not related to the treatment condition. Therefore, we anticipate that student-level attrition will be reasonably low.

Basic formulae for reporting school attrition in each condition, T and C, will be as follows:

- $BaseSch.T$  = number of randomly assigned T schools
- $BaseSch.C$  = number of randomly assigned C schools
- $AssessedSch.T$  = number of T schools in the analytic sample (i.e., those that have not dropped out of the study and for which there are students in the analytic sample with nonmissing outcome data)
- $AssessedSch.C$  = number of C schools in the analytic sample (i.e., those that have not dropped out of the study and for which there are students in the analytic sample with nonmissing outcome data)
- $AttritionSch.T = 1 - (AssessedSch.T / BaseSch.T)$
- $AttritionSch.C = 1 - (AssessedSch.C / BaseSch.C)$
- $Overall\ Attrition\ Sch.\ for\ T\ and\ C = 1 - (AssessedSch.T + AssessedSch.C) / (BaseSch.T + BaseSch.C)$
- $Differential\ Attrition\ Sch.\ for\ T\ vs.\ C = AttritionSch.T - AttritionSch.C$

Overall and differential school attrition will be calculated for each analytic sample.

We will follow the same process for calculating student-level attrition. The Cohort 1 student sample will not include joiners (we will have a roster of students present in study schools in Fall 2015, prior to school random assignment). Sixth-grade students in the study schools in Fall 2015 and prior to school random assignment form the baseline sample. The students from the baseline sample with nonmissing posttests after three years of program implementation will form the analytic sample. Overall and differential student attrition will be calculated in the same way as described above for calculating school attrition; however, student attrition will only be calculated for schools that have not attrited.

### **3.1.6. Cross-overs and no-shows (RCTs only) - optional**

Schools will be analyzed in their originally assigned condition (treatment or control), as long as student achievement data are available. We anticipate no principal cross-overs because control principals will not have access to the EDP training and coaching.

Principal no-shows are a greater concern. It is possible that principals who originally agreed to participate in the EDP and study later withdraw. We will closely monitor implementation data to understand to what extent the study suffers from principal no-shows. We have also included additional, exploratory analyses that examine the impacts of the EDP on student achievement outcomes only for principals who remained in their schools (and in the EDP program for treatment principals) over the course of the three-year period. Moreover, NISL has taken steps to minimize the number of principal no-shows.

### **3.1.7. Data collection for the evaluation of impacts**

*Provide a description of the data collection plan for the confirmatory outcomes and baseline measures.*

The outcome and baseline variables to be used in the confirmatory analyses for this evaluation are individual student scores in mathematics and reading on state-standardized assessments. Student achievement data will include mathematics and reading assessment scores from the *Smarter Balanced*

*Assessment Consortium (SBAC) Exam (California), the Florida Standards Assessment, and for Mississippi, the Partnership for Assessment of Readiness for College and Careers (PARCC) Exam.*<sup>9</sup>

Students’ eighth grade test scores in reading and mathematics will be used as the outcome variables, and students’ fifth grade test scores in reading and mathematics will be used as the pretests. We will monitor each state’s plans in regards to changes in assessments and develop plans to address changes when necessary. Students’ pre- and post-test scores will first be standardized and converted to z scores to combine in the model students who took different assessments.<sup>10</sup> The following table outlines the data collection plan for outcome and baseline measures for all confirmatory and exploratory analyses for the impact study.

Table 2: Outcome Data Collection for the Impact Study

Outcome Measure	Instrument	Administered to /Collected from?	Administered by?	Baseline	Posttest(s)
Student achievement (in reading, mathematics, and science)	SBAC	Students in California study schools	Department of Ed.	Spring 2015	Spring 2018
Student achievement (in reading, mathematics, and science)	PARCC/Questar	Students in Mississippi study schools	Department of Ed.	Spring 2015 (may not be available in all subjects)	Spring 2018
Student achievement (in reading, mathematics, and science)	Florida Standards Assessment	Students in Florida study schools	Department of Ed.	Spring 2015	Spring 2018

<sup>9</sup> In the baseline year, PARCC provided some of the student learning measures in Mississippi, but not all. The state has now decided to eliminate PARCC and switch to Questar for many of its exams in 2016 and beyond.

<sup>10</sup> If the distributions of the assessment scores do not appear to be normal for any assessment, we may explore alternative methods of aggregating results across states including conducting a meta-analysis.

School conditions (multiple outcomes, one per construct)	Teacher CCSR and student CCSR surveys	Teachers and students in California, Florida, and Mississippi in study schools	JHU	Fall 2015 (baseline collected in random subsample of 15 T and 15 C schools in each state)	Spring 2016 Spring 2017 Spring 2018
Teacher retention <sup>11</sup>	Human resource (e.g., school-level teacher retention rate and average years of experience) and school-level student academic performance data	Collected for all teachers and students in study schools from state administrative databases	California, Florida, and Mississippi Department of Ed.	Fall 2015	Fall 2016 Fall 2017 Fall 2018

Other extant data, including school and student covariates and teacher human resource data, will also be collected from state departments of education. For each academic year, we plan to collect school characteristics (e.g., school size, grade level configuration, school type [charter, alternative, magnet], percent female, percent English language learner, percent special education, racial/ethnic composition of students, percent low-income, and school-level achievement), student characteristics (e.g., grade level, gender, race/ethnicity, low-income status, special education status, English language learner status, and student achievement), and teacher human resource data (e.g., school-level teacher retention rate and average years of experience). We will also collect variables to link principals, districts, schools, and students for each academic year.

<sup>11</sup> This exploratory outcome is subject to the availability of the necessary teacher human resource data collected from state administrative databases.

### **3.1.8. Statistical analysis of impacts**

#### **3.1.8.1. Contrasts**

See the attached contrast tool for detailed information about the two confirmatory contrasts that will be estimated in this evaluation. (See Valid36\_Contrast\_Tool.xlsx) We will estimate the impacts of the EDP on mathematics and reading achievement of students in the 8th grade (assessed in Spring 2018) who have been in study schools for three years (i.e., in 6th grade in Fall 2015). We will estimate these two confirmatory contrasts combining data from all three states. In addition, we will estimate a number of exploratory contrasts to address all of the exploratory research questions previously outlined; we did not include the exploratory contrasts in the attached contrast tool.

#### **3.1.8.2. Strategy for dealing with multiple comparisons**

*If the evaluation is not planning to correct for multiple comparisons, this should be stated clearly in this section.*

Where necessary, our plan would be to use an adjusted  $p$ -value, derived from the Benjamini-Hochberg correction of statistically significant effects with multiple comparisons, consistent with WWC standards (WWC, 2014). However, we are not conducting multiple confirmatory tests within the same domain. We are estimating one test of EDP's impact on mathematics achievement and one test of EDP's impact on reading achievement. Because these are tests of impacts in two different domains, a multiple comparisons adjustment is not necessary.

#### **3.1.8.3 Model specifications**

The design reflects a multi-site cluster randomized trial where impact estimates will be calculated using two-level hierarchical linear models (HLMs) with students clustered in schools. For the confirmatory research question, we will examine the impacts of the EDP on students' reading and mathematics achievement in Spring 2018, controlling for baseline achievement (as measured by achievement scores from Spring 2015).

The level-1 (or student) model will include the baseline achievement score (in either reading or mathematics), a set of student covariates (e.g., race/ethnicity, socio-economic status, English language proficiency, special education status, etc.), and a student-level residual. The level-2 (or school) model will include the treatment indicator (treatment or control condition), site (e.g., state and district) blocking variables, a set of school covariates (e.g., principal years of experience, percentage of students who scored proficient or above on state standardized tests, etc.), and a random school-level effect. Separate models will be analyzed for mathematics and reading, resulting in two total confirmatory impact estimates. The HLM model to be used in the student achievement analyses is:

Level-1 (Student) Equation:

$$Assessment\ score_{ij} = \beta_{0j} + \beta_{1j}BaselineScore_{ij} + \beta_{kj} * \sum_2^K StudentCovariates_{ij} + r_{ij}$$

Level-2 (School) Equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Treatment_j + \gamma_{02}California_j + \gamma_{03}Florida_j + \gamma_{0m} * \sum SchoolCovariates_j + \gamma_{0p} * \sum District_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{kj} = \gamma_{k0}$$

where  $\gamma_{00}$  is the covariate-adjusted grand mean assessment score for the control group;  $\gamma_{10}$  is the regression coefficient for the baseline score;  $\gamma_{k0}$  are the vector of regression coefficients for the k student covariates;  $r_{ij}$  is the student residual for student  $i$  in school  $j$ ;  $\gamma_{01}$  is the average treatment effect;  $\gamma_{02}$  and  $\gamma_{03}$  are the regression coefficients for the site indicators (California and Florida, respectively; Mississippi is the reference category);  $\gamma_{0m}$  are the regression coefficients for the m school covariates;  $\gamma_{0p}$  are the regression coefficients for the p districts; and  $u_0$  is the random school effect. All covariates and the baseline score will first be grand-mean centered to facilitate interpretation of the intercept (Enders &

Tofighi, 2007). We plan to use similar models to address the exploratory questions regarding student achievement.

#### **3.1.8.4 Subgroups**

Models similar to the one described above will be used to address the exploratory question regarding the impacts of EDP on student achievement for students in various subgroups. However, the sample will first be restricted to the subgroup of interest (e.g., students living in poverty, ELL students, and students with disabilities).

To address the exploratory question regarding the impacts of EDP on student achievement in schools with higher or lower average student performance, we will include in the model an interaction between the treatment status and a dummy variable indicating whether the school was classified as higher or lower achieving.

#### **3.1.8.5. Decision rules for inclusion/exclusion of covariates**

An a priori decision was made that the pretest measure on which baseline equivalence will be assessed will be included in the model. For other student and school covariates, we will examine the p-values to determine which covariates will be retained in the final model. Student and school covariates whose p-values are less than 0.20 will be retained in the model; covariates with p-values greater than 0.20 will be excluded from the final model.<sup>12</sup>

#### **3.1.8.6. Treatment of missing data.**

Missing outcome data will not be imputed. For categorical student covariates (e.g., socio-economic status), we will create dummy variables indicating that the indicator was unknown if missing. We also do not anticipate any missing data on school covariates because these data are administrative and at the school level. To address students' missing baseline achievement scores, we will explore methods of multiple imputation depending on the proportion of students in the sample who are missing baseline

---

<sup>12</sup> We may use a stricter criterion for inclusion of school-level covariates depending on available degrees of freedom.

achievement scores and the predictive ability of available student covariates that may be used to predict the baseline achievement scores.

If we decide not to impute missing baseline achievement scores and if more than 5% of the student sample is missing baseline achievement scores, we will conduct a sensitivity analysis of the impact effect using the logical bounds methodology discussed by Puma, Olsen, Bell, and Price (2009). This method allows for calculation of best-case and worst-case boundaries of impact when the mechanism for missing data is unknown (when it is unclear if data are missing at random, completely missing at random, or not missing at random). We will use the dummy variable adjustment to impute unbiased values for a share of the missing values for the treatment and control groups and will use the Manski-Horowitz best-case/worst-case scenario for the remainder of the missing values in each group (Puma et al., 2009). Though this approach still encompasses assumptions for some of the missing data (as do all imputation methods), this will provide a boundary of effect sizes which helps to reduce the uncertainty of results that is introduced by missing data. This method can also determine if the impact estimates from the complete-case analyses are trustworthy as the impact estimate should fall within the boundary calculated by this procedure (Puma et al., 2009).

### **3.1.8.7. Calculation of effect size.**

To calculate the effect size, the estimated EDP impact from the analysis model ( $\gamma_{01}$ ) will be divided by the pooled standard deviation derived from student-level sample; the standard deviation is pooled across the treatment and control groups. Impact estimates for each analysis will be reported as effect sizes using Hedges'  $g$ ,

$$g = \frac{M_t - M_c}{\sqrt{\frac{(n_t-1)S_t^2 + (n_c-1)S_c^2}{(n_t + n_c - 2)}}$$

where  $M$  represents the means of the outcome variable for the treatment ( $t$ ) and control ( $c$ ) groups,  $n$  represents the sample sizes for treatment and control groups, and  $S$  is the student-level standard deviations

for the treatment and control groups (WWC, 2014). To indicate statistically significant differences between groups, we will apply a Type-1 error rate (alpha) of 5%.

#### **3.1.8.8. Minimum detectable effects for planned sample.**

The impact study is a cluster randomized-controlled trial where schools are randomly assigned to either the treatment or control condition, within state, and students are nested within schools. The student sample includes sixth grade students in SY 2015-16 in 270 schools in California, Florida, and Mississippi who have non-missing outcome data in the spring of their eighth grade year (in Spring 2018). We assume at least two sixth grade classes per school with 25 students per class, for a total of 50 sixth grade students per school.

To conduct the power analysis and to calculate the minimum detectable effect size (MDES), we made the following assumptions: Type-1 error rate (alpha) of .05, power of .80, intraclass correlation coefficient of .14 for mathematics and .11 for reading (Hedges & Hedberg, 2007), baseline student achievement score that explains 50% of the variance in the outcome achievement score, and school-level covariates that explain 50% of the variance in the school mean achievement. Using PowerUp, a power analysis tool developed by Dong and Maynard (2013), we obtained an MDES of less than 0.10 standard deviations for mathematics and reading achievement. Assuming an attrition rate of 30% at the student level and 10% at the school level over the three-year period, the MDES remains less than 0.10 standard deviations for mathematics and reading achievement.

#### **3.1.9 Baseline equivalence testing – QEDs and RCTs with high attrition**

WWC standards do not require that baseline equivalence be established if attrition is low, but evidence of baseline equivalence will be required if cluster (school) or student attrition is high. If attrition is high at either the cluster or student level, treatment and control students in the analytic sample will be considered equivalent at baseline if the standardized mean difference in the baseline score is less than 0.25 standard deviations, provided that the baseline achievement score is included as a covariate in the analytic model. We will assess the baseline equivalence of students in the analytic sample using the same HLM approach as outlined earlier in this document. In this case, the outcome variable will be the baseline

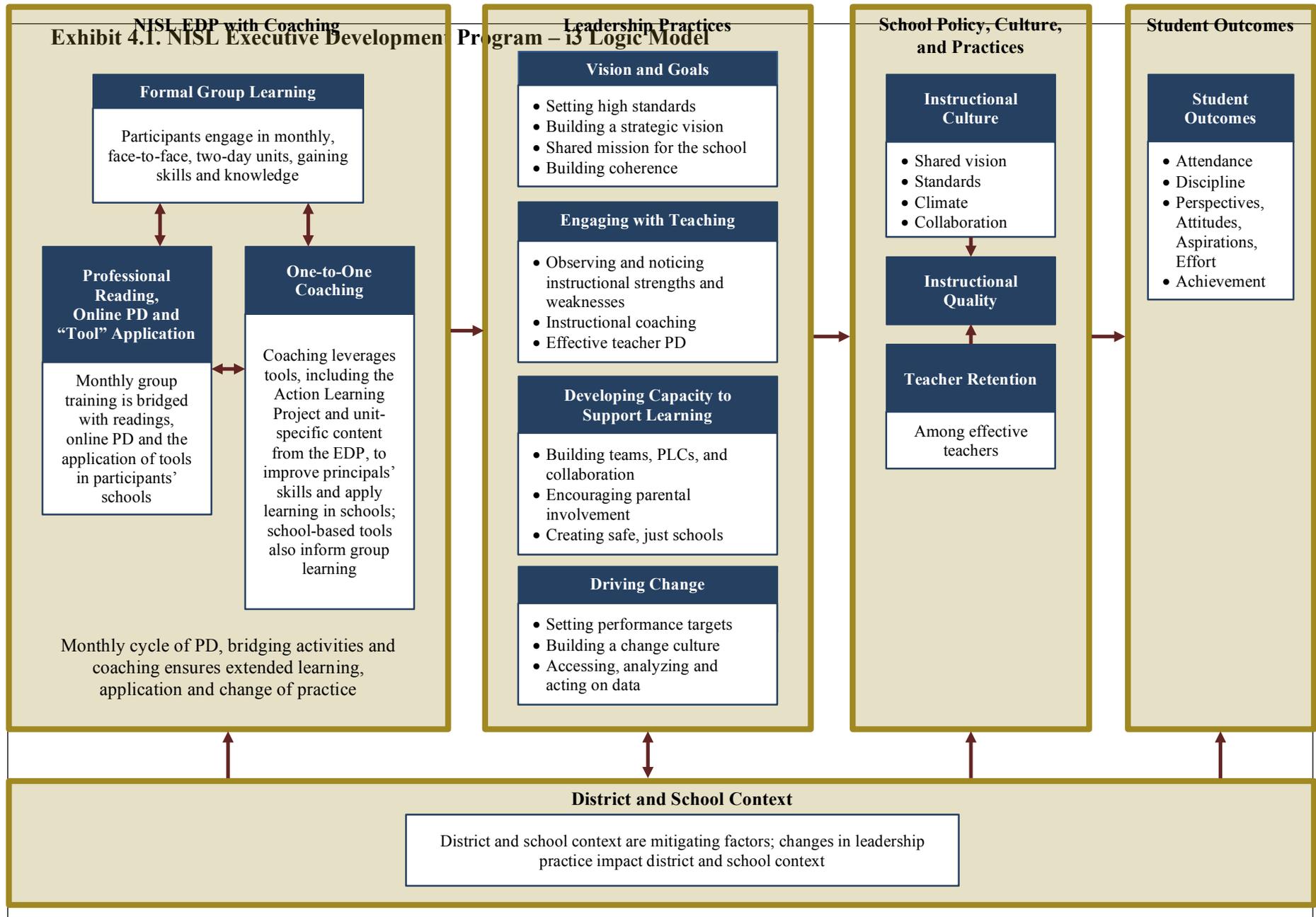
reading or mathematic achievement score from Spring 2015, and the independent variables will be the treatment indicator and site indicators.

## 9. Implementation Evaluation

### 3.2. *Logic model for the intervention*

The figure below (Exhibit 4.1) outlines the theory of change proposed by the NISL EDP program. As demonstrated in this figure, the combination of NISL EDP curriculum and NISL coaching leads to the refinement of principal leadership practices, which in turn leads to changes with school policy, culture, and practices within the principal's school. Finally, as a prospective result of these school-based changes student outcomes are improved.

**NISL EDP with Coaching**  
**Exhibit 4.1. NISL Executive Development Program – 15 Logic Model**



### **3.3. Research questions for evaluation of implementation**

#### **Fidelity of Implementation (FOI) Questions**

1. Across all states, is each of the key components of the EDP intervention implemented with fidelity across the full sample of schools?
2. For each state site, what percent of NISL schools have a high FOI according to the FOI matrix?

#### **Program Quality Questions**

3. How do principals and coaches perceive the effectiveness of the EDP? How has EDP changed principal priorities, practice, and how they allocate their time?
4. How do principals perceive the quality of the support they received through coaching?

### **3.4. Measuring fidelity of implementation**

JHU will focus on objective measures of fidelity of implementation (FOI) using a FOI matrix that is designed to judge adherence to the program model. To determine whether the EDP program and coaching were implemented with fidelity, we will measure fidelity to the three key components of EDP (as shown in the logic model): one-on-one coaching, formal group learning, and professional reading. The FOI will be rated using criteria that represent high delivery and participation at the school/principal levels as well as for the full sample of schools/principals. The FOI evaluation will examine the extent to which each component of the EDP was implemented as intended for the treatment group of principals only. See pages 48-49 for the FOI matrix.

In Fall 2015 (Year 1), all 135 EDP treatment principals will begin participating in the EDP program and their participation will extend to Fall 2016 (Year 2). The predominant training (and implementation) activity consists of approximately 24 classroom days of professional development that EDP principals attend in regional locations. Participation is directly recorded via attendance logs. NISL

also will provide coaching to the treatment group principals from Spring 2016 to Summer 2017.

Delivery of and participation in coaching activities will be measured via coach and coachee logs.

Finally, delivery of related curricula materials to treatment principals for professional reading will be captured by shipping receipts.

The attached FOI matrix (pages 48-49) details how the FOI will be scored. Each key component will be measured based on one indicator, which will be scored at the principal/school level. For each indicator, a threshold for implementation with fidelity at the principal/school level is defined. An aggregate (i.e., sample-level) fidelity score will be computed for each indicator/key component by determining the overall percentage of schools with high implementation. A threshold for implementation with fidelity at the sample level is defined in the fidelity matrix for each of the three key components. Fidelity of implementation will be measured overall for all three states, and also separately for each state.

The FOI for each component will be measured according to the framework annually for each school year that the intervention is implemented—SY 2015–2016 and SY 2016–2017.<sup>13</sup> In other words, implementation fidelity will be measured twice during the grant period. Each year, fidelity will be measured for the full sample of principals/schools assigned to the treatment group—approximately 135 principals across all three states.

A second feature of the implementation study will involve gauging the perceived quality and effectiveness of the EDP. This focus will be addressed by a number of data collections (see the next section for more detail). In addition, JHU will conduct case studies in six schools (two per state) to shed light on how EDP novice principals use their new knowledge and skills to change key aspects of how their schools function to increase student achievement. These schools will be selected from the population of treatment schools. For each case study school, JHU will visit the school three times over the course of the study (Fall 2016, Fall 2017, and Fall 2018), for a total of 18 visits for all six case

---

<sup>13</sup> The EDP training and coaching occur for only two years. The main confirmatory contrast focuses on student achievement outcomes in the third year of the study when fully-trained principals are leading schools.

study schools. In addition to the data collection measures used for all treatment and control schools, additional data on principals in case study schools will be collected by the research team using the “Principal Shadowing Tool” and principal interviews.

Analyses of qualitative data will be guided by Miles and Huberman’s (2004) model, consisting of transcribing the responses, deriving codes, identifying themes, and revising and refining based on member checking and interrater review. Triangulation across data sources and methods will be used to validate the major findings. Cross-case analyses, member-checked through an online focus group of key informants (Miles & Huberman, 2004), will be used to identify programmatic, individual, and contextual factors. Each case study will be member-checked with key informants (e.g., novice middle school principals and teacher leaders). Coding for quality assurance will be applied to the analyses of qualitative data.

Specifically, CRRE analysts will subject the qualitative data sources to coding using NVivo, a software program that facilitates coding, analysis, and reanalysis of qualitative data in multiple file formats. NVivo is a system created for mixed-methods research that allows multiple users to collaborate by viewing and accessing qualitative and quantitative data within an Internet-based system. NVivo enables researchers to upload files for coding, store, and manage data securely, link qualitative and quantitative data, and identify data patterns for further analyses. Two principles will be utilized to analyze these data: (a) triangulation and (b) grounded theory.

In triangulation, the results from each of the data sources will be compared against one another to give a more comprehensive view. It is “a nonexperimental qualitative sociological method that employs an exhaustive examination of cases in order to prove universal, causal generalizations” (Vidich & Lyman 1994, p. 39). Grounded theory is one way in which the relationship between theory and data in this research will be formulated. Grounded theory has been defined as “theory that was derived from data, systematically gathered and analyzed through the research process. In this method, data collection, analysis, and eventual theory stand in close relationship to one another” (Strauss &

Corbin 1998, p. 12). A key process in grounded theory is coding, which entails processes whereby data are broken down into component parts and given labels (names). Strauss and Corbin (1990) have distinguished three types of coding practice, all of which will be used in this research. They are as follows:

**Open coding:** “The process of breaking down, examining, comparing, conceptualizing and categorizing data” (Strauss & Corbin, 1990, p. 61); this process yields concepts, which are later to be grouped and turned into categories.

**Axial coding:** “A set of procedures whereby data are put back together in new ways after open coding, by making connections between categories” (Strauss & Corbin, 1990, p. 96). This procedure is undertaken by linking codes to contexts, consequences, patterns of interaction, and causes.

**Selective coding:** “The procedure of selecting the core category, systematically relating it to other categories, validating those relationships, and filling in categories that need further refinement and development” (Strauss & Corbin, 1990, p. 116). A core category is the central issue or focus around which all other categories are integrated. It is what Strauss and Corbin have called the storyline that frames your account (cited in Bryman, 2008, p. 543).

Coding the data in this way typically leads to a number of outcomes (for further discussion, see Bryman, 2008, p. 544), which are as follows: concepts, category, hypotheses (initial hunches about relationships between concepts), and theory (a set of well-developed categories that form a theoretical framework).

### **3.5. *Data collection plan***

Specific measures that will be collected as part of the implementation study are listed below. First, we describe data that will be collected to score FOI, which will be scored as described in the fidelity matrix (pages 48-49). These data primarily reflect program delivery and principal participation

---

in the EDP training and coaching. Second, we describe data that will be collected to measure the perceived quality and effectiveness of the EDP program and coaching.

**Data to Measure FOI:**

*Coach and coachee logs:* Coaches will both be required by NISL to complete short logs after each month of coaching, and coachees will be required to complete short logs after each quarter. The quarterly coachee logs require coachees to report on the amount of time spent in coaching sessions, the type of coaching (in-person, phone, etc.), and focus of the coaching. Coachees are also asked to reflect on the effectiveness and benefits of coaching sessions, and in particular, to provide further detail on the content of coaching sessions and resulting changes in practice. Logs are submitted to the State Coordinators, who will disseminate them to NISL staff. Coaches will also complete regular coaching logs on the type, content, and results of coaching, and submit the logs to State Coordinators quarterly for compiling and sharing of the data.

*Attendance logs:* Attendance records from each day of the two-day training units held in regional locations will be obtained to measure the number of days each principal attends monthly, face-to-face formal group learning.

*Shipping company receipts:* Receipts from the shipping company will be used to document whether or not each principal received the full set of training materials for Phase I and Phase II training. These receipts will provide evidence that the professional reading component of the EDP program was delivered to participants and implemented with fidelity. Each principal should receive materials at least twice—at the start of SY15-16 and in the summer before SY16-17—both prior to and during the EDP.

**Data to Understand Perceived Program Quality and Effectiveness:**

*Training surveys.* Each training session will be evaluated by NISL using a brief survey asking participants to rate clarity, organization, and relevance in relation to the logic model's direct outcome

---

---

dimensions. These surveys, which include Likert-type ratings items and open-ended questions, are used routinely by NISL in its EDP training. The surveys will be administered online by NISL.

*Training observations.* JHU will draw a purposeful sample of nine training sessions (three per state) to be observed. In each site, JHU will record impressions using an informal “observation guide” designed to capture (through open and prompted field notes) key dimensions of the training (e.g., organization, clarity, and time management), adaptation to diversity (school characteristics, principal experiences), relevance to session goals, and participant engagement. This observation guide will be developed by CRRE in collaboration with NISL.

*Principal focus groups.* At the completion of the EDP training and again at the completion of coaching, six focus groups (two per state, each consisting of six-to-eight randomly selected principals) will be conducted by webinar to determine reactions to the EDP implementation (i.e., curriculum content, professional development activities, and coaching) in relation to the logic model outcomes.

*Coach focus groups.* Three focus groups (one per state) of randomly selected coaches will be conducted by conference call at the completion of coaching. Questions will focus on principals’ participation, their interest in such guidance, the ability to capitalize on it, their readiness to take risks, and their ability to capitalize on what the EDP offers them.

*NISL principal survey.* At the end of each school year, principal participants will be asked to complete a survey regarding their experiences in the EDP program and its applications to leadership activities in their schools (for treatment principals only). Treatment and control principals will also be surveyed about the actions they have taken to improve the quality of education students are receiving and to improve the working environment and culture for teachers and students. Survey data will also be used to determine and monitor the diffusion of treatment to principals in the control condition.

*Principal shadowing.* For the purposes of the study, an “EDP Principal Shadowing Tool” will be developed by CRRE in consultation with NISL for use in the six case study schools. The tool will be designed to collect reliable and rich qualitative data on principals’ everyday practices relative to

and separate from their EDP training. It will involve a guided conversation and reflections between the researcher and the principal as the latter goes about a typical school day. Researchers will also interview principals at the six case study schools. The following table outlines the data collection timeline for the implementation study.

Table 3: Data Collection for the Implementation Study

Data Collection Effort	Sample	Administered by?	Timing
Coach and coachee logs	Coaches and principals in treatment schools	NISL	Once a month for coaches and once a quarter for coachees during coaching
Attendance logs	Principals in treatment schools	NISL	After in-person formal group learning sessions from 9/15–11/16
Shipping receipts	Principals in treatment schools	NISL	Ongoing from 9/15–11/16
Training surveys	Principals in treatment schools	NISL	Ongoing from 9/15–11/16 after each training session
Training observations of session quality	A purposeful sample of nine training sessions (three per state) will be observed and impressions recorded on an “observation guide” that will be designed in collaboration with NISL	CRRE	Ongoing from 9/15–11/16 after select training sessions
Principal focus groups	Six focus groups (two per state, each consisting of six-to-eight randomly selected principals) will be conducted by webinar to determine reactions to the EDP implementation	CRRE	Winter 2017 and Spring 2018
Coach focus groups	Three focus groups (one per state) of randomly selected coaches will be conducted by conference call	CRRE	Summer or Fall 2017
NISL principal survey	Principals in control and treatment schools	NISL	Spring 2016, Spring 2017, and Spring 2018

Principal shadowing (and interviews)	Principals in the six case study (treatment) schools	NISL	Fall 2016, Fall 2017, and Fall 2018
--------------------------------------	--	------	-------------------------------------

### **3.6. Fidelity reporting plan**

Table 4 presents an example illustrating how the FOI may be reported. Each key component will be measured based on one indicator, which will be scored at the principal/school level. For each indicator, a threshold for implementation with fidelity at the principal/school level is defined. An aggregate (i.e., sample-level) fidelity score will be computed for each indicator/key component by determining the overall percentage of schools with high implementation. A threshold for implementation with fidelity at the sample level (i.e., pooled across all three states) is defined in the fidelity matrix for each of the three key components. Fidelity of implementation also will be measured separately in each state.

The FOI for each component will be measured according to the framework (pages 48-49) annually for each school year that the intervention is implemented—SY 2015–2016 and SY 2016–2017. In other words, implementation fidelity will be measured twice during the grant period. Each year, fidelity will be measured for the full sample of principals/schools assigned to the treatment group—approximately 135 principals.

Table 4: Example Table for Reporting Fidelity Findings

Intervention Components	Number of measurable indicators representing each component	Sample	Component level threshold for each sample-level unit	Evaluator's criteria for "Implemented with Fidelity"	Percentage of schools/principals where component was "Implemented with Fidelity"	Implemented with fidelity? (Yes, No, N/A)
One-to-one coaching	1	135 principals <sup>1</sup>	Received at least 90% of in-person coaching visits	At least 90% of principals met threshold	TBD	TBD
Formal group learning	1	135 principals	Attended at least 90% of group learning days	At least 90% of principals met threshold	TBD	TBD
Professional reading, online PD, and "tool" application	1	135 principals	Obtained full set of training materials	At least 90% of principals met threshold	TBD	TBD

\*TBD = To be determined. This information will be available after implementation fidelity data are collected and scored.

<sup>1</sup> The sample includes 52 principals in California, 43 principals in Florida, and 40 principals in Mississippi.

Exhibit 4.2. Valid36 NISL Fidelity of Implementation (FOI) Matrix

Indicators	Definition	Unit of implementation	Data Source(s)	Data Collection (who, when)	Score for levels of implementation at unit level	Threshold for implementation with fidelity at unit level	Roll-up to program level (score and threshold for implementation with fidelity at sample level)	Expected sample for fidelity measure	Expected years of fidelity measurement
<b>Key Component 1: Formal Group Learning</b>									
<b>Indicator 1</b>	Monthly, face-to-face 2-day units  Each principal participates for a total of 24-classroom days (12 2-day units) typically over 12-13 months – approx.. 16 in SY15-16 and 8 in SY16-17	Principal/school	Attendance Logs	NISL provides JHU with Attendance Logs bimonthly	Percentage of total classroom days  0= <80% of total classroom days  1= 80-89% of total classroom days  2= 90-100% of total classroom days	Implementation with fidelity =  Score of 2	Percentage of principals with a score of 2  0 = <90% of principals  1 = at least 90% of principals  Implementation with fidelity = score of 1		
<b>All indicators</b>							<i>Indicator 1 score is the component score</i> Range = 0-1  Fidelity =1	All principals across all three states (approximately 135 principals)	SY 2015-2016 SY 2016-2017
<b>Key Component 2: One-to-One Coaching</b>									
<b>Indicator 1</b>	One-to-one (coach-to-principal) communication, via in person visits, electronic, oral, and written means plus at least 1 full day in-person visit at the outset	Principal/school	Coach and Coachee Reports	NISL provides Reports to JHU bimonthly	Percentage of total coaching days received  0 = <80% of days  1= 80-89% of days  2= 90-100% of days	Implementation with fidelity =  Score of 2	Percentage of principals with a score of 2  0 = <90% of principals  1 = at least 90% of principals  Implementation with fidelity = score of 1		

	Each principal receives a total of 11 days of communication – 3 in SY15-16 8 in SY16-17								
<b>All indicators</b>							Indicator 1 score is the component score Range = 0-1  Fidelity =1	All principals (approximately 135 principals)	SY 2015-2016 SY 2016-2017
<b>Key Component 3: Professional Reading, Online PD, and “Tool” Application</b>									
<b>Indicator 1</b>	Each principal receives a full set of materials for the school year	Principal/school	Shipping company receipts	NISL provides shipping confirmation to JHU	0 = No materials 1 = Incomplete set 2 = Full set	Implementation with fidelity =  Score of 2	Percentage of principals with a score of 2  0 = <90% of principals  1 = at least 90% of principals  Implementation with fidelity = score of 1		
<b>All indicators</b>							Indicator 1 score is the component score Range = 0-1  Fidelity =1	All principals (approximately 135 principals)	SY 2015-2016 SY 2016-2017

## References

- Bryk, A.S., Sebring, P.B., Allensworth, E., Luppescu, S., & Easton, J.Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: Consortium on Chicago School Research.
- Bryman, A. (2008). *Social research methods* (3rd ed.). New York, NY: Oxford University Press.
- Cheney, G.R., Davis, J., Garrett, K., & Holleran, J. (2010). *A new approach to principal preparation: Innovative programs share their practices and lessons learned*. Fort Worth, TX: Rainwater Leadership Alliance.
- Consortium on Chicago School Research. (CCSR). (2015). *My Voice, My School Survey: Student measures reliability data*. Chicago, IL: Consortium on Chicago School Research.
- Council of Chief State School Officers. (2008). *Interstate School Leaders Licensure Consortium (ISLLC) standards*. Washington, DC: CCSSO.
- Dong, N. & Maynard, R. A. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, 6(1), 24-67. doi: 10.1080/19345747.2012.673143
- Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, Vol 12(2), 121-138. { HYPERLINK "http://psycnet.apa.org/doi/10.1037/1082-989X.12.2.121" \t "\_blank" }
- Hedges, L. & Hedberg, E. (2007). Intraclass correlation values for planning group randomized trials in education. *Educational Evaluation and Policy Analysis*, 29 (1), 60-87.
- Levenstein, R. (2015). *Survey information for the 2015 CPS Teacher Survey*. Chicago, IL: Consortium on Chicago School Research.
- The Meristem Group. (2009). *National Institute for School Leadership (NISL): Massachusetts program implementation 2005-2008*. Boston, MA: Meristem Group.
- Miles, M. B., & Huberman, A. M. (2004). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Nunnery, J. A., Ross, S. M., Chappell, S., Pribesh, S., & Hoag-Carhart, E. (2011). *The impact of the NISL Executive Development Program on school performance in Massachusetts: cohort 2 results*. Norfolk, VA: The Center for Educational Partnerships at Old Dominion University.
- Nunnery, J. A., Yen, C., & Ross, S. M. (2011). *Effects of the National Institute for School Leadership's Executive Development Program on school performance in Pennsylvania: 2006-2010 pilot cohort results*. Norfolk, VA: The Center for Educational Partnerships at Old Dominion University.
- Public Impact. (2008). *School turnaround leaders: Competencies for success*. Chapel Hill, NC: Public Impact for the Chicago Public Education Fund.
- Puma, M., Olsen, R., Bell, S., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

- Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods (advanced quantitative techniques in the social sciences)* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S.W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group randomized experiments. *Educational Evaluation and Policy Analysis*, 29 (1).
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.
- Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.
- University of Chicago. (2015a). *My Voice, My School Student Survey*. Chicago, IL: Consortium on Chicago School Research.
- University of Chicago. (2015b). *My Voice, My School Teacher Survey*. Chicago, IL: Consortium on Chicago School Research.
- Vidich, A. J., & Lyman, S. M. (1994). Qualitative methods: Their history in sociology and anthropology. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 23–44). Thousand Oaks, CA: Sage.
- What Works Clearinghouse. (2014). *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0)*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.