CAPE Reference No. 2006-001-01
MPR Reference No.: 6319-520
Joyce Grant No: 06-30021

MATHEMATICA
Policy Research, Inc.

Evaluation of the Teacher Advancement Program (TAP) in the Chicago Public Schools: Study Design Report

November 14, 2007

Steven Glazerman
Allison McKie
Nancy Carey
Dominic Harris

Prepared for:

The Joyce Foundation
70 West Madison Street, Suite 2750
Chicago, IL 60602-4317
Telephone: (312) 782-2464
Facsimile: (312) 782-4160

Program Officer:
Gretchen Crosby-Sims

Prepared by:

Mathematica Policy Research, Inc.
600 Maryland Ave. S.W., Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Principal Investigator:
Steven Glazerman

CONTENTS

Chapter **Page**

TABLES

FIGURES

CHAPTER I

BACKGROUND

Recent evidence has confirmed that teacher quality is a critical component in student achievement. Research indicates that teachers have a distinct and lasting effect on learning, with teacher quality outweighing other school-based factors such as racial composition or class size (McCaffrey et al. 2003; Sanders and Horn 1998). The relative effect of a high-quality teacher in the classroom is potentially large and accumulates over time for students consistently taught by effective teachers (Hanushek and Rivkin 2003; Sanders and Rivers 1996).

While researchers agree on the importance of teacher quality for student outcomes, the scientific and policy communities have been struggling to determine the best way to measure teacher effectiveness and to recruit and retain the teachers identified as effective. The most obvious policy response is to reform how we evaluate and compensate teachers. The traditional system, known as the uniform salary schedule, pays teachers based solely on years of experience and educational attainment, and does not incorporate formal evaluation in compensation determinations. Many efforts to reform the uniform salary schedule, including performance-based pay (merit pay), responsibility pay (career ladders), and bonuses for teaching in high-need subjects or geographical areas, have been largely unsuccessful, typically failing to survive even long enough to be studied (Glazerman 2004). Many authors have examined these failures of implementation and their possible explanations (Murnane and Cohen 1986; Hatry et al. 1994; Podgursky 2002; Glazerman 2004). There are many possible reasons for these failures, but not least among them are the challenges of developing objective and reliable measures of teacher performance and the lack of convincing research evidence demonstrating that teacher pay reforms—which typically involve pay increases—are worth the extra expense.

Recent developments, however, have altered the conditions such that a breakthrough is very likely. The data and technology for measuring teacher performance have grown dramatically in recent years as a result of the increased frequency of student testing under the No Child Left Behind Act, the growth of more sophisticated student data tracking systems, and the advancement of methods for computing value added indicators of teacher

performance.[1] These trends have opened up new program and research opportunities (Glazerman et al. 2006). At the same time, political and financial support for teacher compensation reform has grown, as demonstrated by the large federal investment in the U.S. Department of Education's Teacher Incentive Fund (TIF) ($95 million in its first year); by state initiatives such as Minnesota's Q-Comp program ($86 million) and Nevada's teacher bonus programs ($65 million); and by local initiatives such as New York's Partnership for Excellence ($15 million granted by the Carrol and Milton Petrie Foundation). Twenty governors described teacher compensation as a major education issue in their 2005 State of the State addresses, and nine specifically mentioned performance-based or merit pay (Azordegan et al. 2005). Recent policy papers promoting reform of the teacher pay system include a Progressive Policy Institute report that argues for simultaneously raising teacher pay and reforming the teacher pay system (Hassell 2002); a report by The Hamilton Project, an initiative of the Brookings Institution, that recommends offering highly effective teachers bonuses for teaching in high-poverty schools (Gordon et al. 2006); and two sets of guidelines on reforming teacher pay by teacher research and advocacy organizations, one called TeacherSolutions (2007), a group of 18 nationally recognized teachers, and the other called the Working Group on Teacher Quality (2007), which includes several teacher-related think tanks.

One program that is poised to take advantage of these favorable conditions is the Teacher Advancement Program (TAP), developed by the Milken Family Foundation in the late 1990s. TAP, described in detail below, is a whole-school approach to evaluating and compensating teachers and providing professional development opportunities to both improve teaching and help schools attract and retain good teachers. The program, which includes value added assessment of teacher performance, professional development, career ladder opportunities, and performance-based bonuses, has been adopted in over 100 schools across a dozen states to date.

Yet some policymakers have expressed concerns about the lack of rigorous, independent research on the effectiveness of TAP or any other approach to teacher compensation reform (Hassell 2002). Given the pace of policy proposals and investment in new teacher pay reforms, the research needed to guide these investments is lagging. To date, the research literature consists of no experimental studies[2] and very few quasi-experimental studies—including TAP studies by Schacter et al. (2002 and 2004) and Solmon et al. (2007), and a study by Clotfelter et al. (2006) of North Carolina's $1,800 retention bonus program. The three TAP studies were conducted by the developers of the program, and two of them relied on small, self-selected comparison groups of schools in two states. The more recent

[1] "Value added" is the term frequently used to refer to teachers' contribution to student achievement, or their productivity in bringing about student learning gains.

[2] An experimental study is a randomized controlled trial—the gold standard for generating causal evidence—in which study subjects are randomly assigned to treatment or control group status so that outcome differences between them can be attributed to the treatment. Quasi-experimental or non-experimental evaluation studies rely on methods that attempt to approximate random assignment, but require additional assumptions before causal inferences can be drawn (Cook and Campbell 1979; Myers and Dynarski 2003).

report includes larger numbers of comparison schools and teachers, a total of 61 TAP and 285 non-TAP schools across six states. The North Carolina study acknowledged flaws in the program's implementation, specifically that teachers who were subject to the bonus were often unaware of their eligibility until after they received a payment.

To address this lack of research evidence, the Chicago Public Schools (CPS) has teamed with an independent third-party evaluator, Mathematica Policy Research, Inc. (MPR), to help design and carry out a five-year demonstration that will provide a rigorous test of the impacts of TAP in high-need schools. The CPS effort, called the Recognizing Excellence in Academic Leadership program (Project REAL), uses the TAP program model and includes incentive pay for school principals. CPS receives funding for the REAL program from the U.S. Department of Education through a Teacher Incentive Fund grant. To maintain the objectivity of the evaluation of Project REAL, the research component is funded by a separate grant from the Joyce Foundation, and MPR's work is entirely independent of Project REAL and the Chicago Public Schools.

The evaluation is a five-year effort. Study Year 1 is for planning and recruiting. In Spring 2007, we randomly assigned eligible schools to implement the program in Year 2 or Year 3 according to a lottery that is discussed below. In 2008, we conduct teacher and principal surveys and collect administrative records data from CPS on teachers and schools in TAP and non-TAP schools. We continue this data collection in the remaining three years and conduct another round of random assignment in 2009 (for another set of schools to begin implementing TAP in Year 4 or Year 5) and another round of teacher and principal surveys. The study will produce annual reports and a final report in Year 5. The study team will brief key stakeholders on findings as they become available. The rest of this report provides more background on the study and lays out the study design in more detail.

DESCRIPTION OF TAP IN CHICAGO

The Teacher Advancement Program was developed by the Milken Foundation in the late 1990s with the aim of helping schools put a highly skilled, strongly motivated, and competitively compensated teacher in every classroom in America. According to its developers, TAP is a comprehensive, research-based school reform that seeks to attract talented people to the profession and create an environment in which they can thrive. It does so by offering sustained opportunities for career advancement and ongoing school-based professional development, by insisting on instructionally focused accountability, and by providing performance pay.

Program Components

Schools implement the program, with leadership and guidance from its developer, the National Institute for Excellence in Teaching (NIET). TAP can vary somewhat from school to school, but must contain the following four elements:

Multiple career paths give qualified teachers opportunities to take on more responsibility and receive compensation for doing so. Teachers in a TAP school can be career teachers, mentor teachers, or master teachers. Along with master and mentor teacher status come more responsibility and more pay.

Ongoing applied professional growth provides teachers with school-based professional development during the school day. Teachers meet weekly in small "cluster" groups led by a master teacher, and together analyze student data, improve instruction, and learn new research-based instructional strategies that increase their students' academic achievement.

Instructionally focused accountability ties teacher evaluations to teaching skills and student achievement. A team of classroom observers who are trained and certified by NIET to rate teaching skills according to clearly defined criteria conduct four to six observations during the year. CPS, in conjunction with a team of experts at the University of Wisconsin's Center for Education Research, will use test score data to estimate value added indicators of teacher performance based on student achievement gains.

Performance-based compensation provides bonuses to teachers who demonstrate their skills through classroom evaluations and who increase their students' academic growth over the course of the year. TAP's professional development is designed to support teachers in achieving these goals. The bonuses in Chicago are expected to be drawn from a pool that averages $4,000 per teacher per year, allocated according to a formula that places equal weight on classroom observation and value added measures. School principals will also be eligible for bonuses based on similar criteria, plus a measure of the degree of TAP implementation in the school.

The Chicago TAP model implemented under the REAL program will have several specific locally tailored attributes, which have been negotiated with key stakeholders including the district leadership, Project REAL staff, NIET, and the Chicago Teachers Union (CTU). These locally specific attributes include the rules pertaining to the staff who are covered by the incentive plan, the types and amount of training provided by NIET, the rules for posting and filling mentor and master teacher positions, the precise method for evaluating and quantifying teachers' performance, and the size of the performance and career ladder bonuses. In a future report we will describe the Chicago TAP model in detail and contrast it with other models that are in use around the country.

Program Logic Model

Apart from the particular program rules and attributes, we have developed a simple logic model that helps us understand the mechanisms by which TAP is expected to produce impacts. The first mechanism is what we refer to as a teacher *productivity effect*. We hypothesize that teachers in schools implementing TAP will be motivated through incentives and enabled through professional development and formative evaluation to improve their practice, particularly as it relates to producing student achievement growth. We also hypothesize that students in those schools may benefit from an overall climate of increased

professionalism and satisfaction on the part of all staff who are included in the bonus pool. However, we also allow for the possibility that there are unintended negative consequences of TAP. For example, a fear raised by critics of incentive programs is that the existence of incentives and within-school promotion opportunities could create competition and discourage collaboration among teachers.

Another mechanism by which TAP can affect student outcomes is through a *composition effect*, which means that TAP is expected to change the mix of teachers to favor more talented individuals. The teacher composition effect is a combination of recruitment and retention effects; that is, if TAP is successful in rewarding effective teachers, then ineffective teachers would prefer to leave the school and more effective teachers would prefer to stay on longer than they would otherwise. Similarly, a school program that rewards effective teaching should be more attractive to teacher candidates who are likely to perform well on the teacher evaluations and value added measures, allowing the school to attract a more talented pool of candidates than it would otherwise. Even if the teacher bonuses were undifferentiated and resulted in everyone in the school receiving higher pay, it should attract more candidates, all other things being equal.

An important dimension of expected TAP effectiveness in Chicago is the maturity of the program. By necessity, we will focus on TAP in its early years in Chicago, not in the steady state. Impacts during the startup phase may be unrepresentative of steady state impacts for at least two reasons. First, the startup effects could be positive in that increased attention and scrutiny could result in more resources being trained on the early adopters (schools) than would be possible under wider implementation. Conversely, the startup effects could be negative, in that the early adopters have less time than later adopters to learn about the program, understand its rules, and hire staff who could best take advantage of it.

We will attempt to model two types of startup effects: one for startup in a *school* and one for overall program maturation in the *district*. School maturation occurs because the second year of implementation may differ from the first and the third from the second, and so on, because the principal and staff have more time to adjust and learn. The school-based team that evaluates teachers will have more practice. The teachers in the school will have had more time to understand and change their teaching in response to the incentive system. Program maturation in the district occurs because CPS will learn over time how to select schools to participate in TAP and how to train the school's staff, and will have opportunities to adjust program rules and refine methods for calculating teachers' value added, for example, as they gain experience with the model.

Finally, we hypothesize that there could be heterogeneous treatment effects. Regardless of whether the program has beneficial effects, deleterious effects, or a combination of both overall, the impacts may vary by categories of teachers. For example, teachers in non-tested grades and subjects will be evaluated differently and perhaps affected differently than those whose own students' test scores play a role in their performance evaluation. Experienced teachers may be affected differently than new teachers or those with different types of training—for example, early career teachers may be on a steeper part of their learning curve and therefore have greater capacity to benefit from frequent evaluation and feedback.

However, less experienced teachers may have weaker incentives than their more experienced counterparts if they are not in as good a position to compete for mentor and master teacher roles within the school.

RESEARCH QUESTIONS FOR THE EVALUATION

The main goal of the evaluation is to estimate the impacts of TAP on teacher quality in terms of student achievement and the recruitment and retention of teachers. We will also study the program's implementation and compare it with prevailing practices to help us understand the impact findings and put them in context. The impact and implementation analysis will each be guided by a set of research questions:

Impact Questions

1. What is the impact of TAP on teachers' "value added" to student achievement and on career decisions (transfers and retention)?

2. How much of the TAP impact on teacher quality is a productivity effect (teachers doing their job better) versus a composition effect (improving the mix of teachers)?

3. Does the program become more effective over time as the school has more experience adapting to it?

4. Does the program have greater impacts for some types of schools or teachers than others?

Implementation Questions

1. Was TAP implemented faithfully according to the model outlined by its developers?

2. What factors facilitate or impede the successful implementation of TAP?

3. How does TAP differ from what is normally implemented in CPS schools? Specifically, how prevalent are professional development and mentoring, teacher-leader roles, observation of teacher practice, and feedback in CPS schools?

PLAN OF THIS DESIGN REPORT

The rest of this report lays out the study design in more detail. In Chapter II, we provide an overview of the technical approach, including random assignment to year of implementation and a comparison school design. Chapter III discusses data collection and maps the data collection onto the research questions listed above. Chapter IV discusses analysis and reporting.

CHAPTER II

OVERVIEW OF THE STUDY DESIGN

We have designed the TAP evaluation to allow us to describe the context, program implementation, and outcomes of TAP schools in Chicago and compare them to TAP schools elsewhere and to comparable non-TAP schools in Chicago. We place particular emphasis on the contrast between TAP and non-TAP schools. To create the most rigorous study possible that is flexible enough to answer a variety of research questions, we will use a hybrid study design that relies on both experimental methods (random assignment) and quasi-experimental methods (propensity score matching) to identify a set of students, teachers, and schools that can serve as a local comparison group. Described in more detail below, these two approaches help us to approximate the counterfactual state—what *would have happened* in the absence of TAP. We will use data from NIET on TAP schools, and from CPS on student test scores and teacher background/mobility for all schools in the district. We will also collect primary data through surveys of teachers and principals conducted at strategic points during the study to measure teacher and school background, context, program implementation, and outcomes. In addition, we will gather program documents and interview key program officials to inform the descriptive and implementation analyses.

For the impact analysis, we will compare the outcomes in the TAP ("treatment") schools with those of the non-TAP ("control" or "comparison") schools. Similarly, for the implementation analysis, we will compare TAP and non-TAP schools on the degree to which staff engage in activities that comprise the TAP model and those that may not be TAP-specific but are related to teacher quality. For example, how are teachers in TAP schools evaluated, how often, and by whom? How does this compare with evaluation of teachers in non-TAP schools? What is the form of feedback and professional development? In addition to the comparative analysis of implementation in TAP and non-TAP schools, we will incorporate data from NIET on the ratings of fidelity of TAP schools in Chicago to the TAP model as adopted elsewhere around in the country. The rest of this chapter describes the design in more detail.

IMPACT ANALYSIS

The impact analysis will focus on several outcomes, the primary one being student achievement. Another important outcome is teacher retention. For retention outcomes, we care about not only the percentage of teachers retained, but also their qualifications and quality. Chapter III discusses our approach to measuring these attributes. Finally, we will estimate impacts on intermediate outcomes, such as teacher attitudes and program experiences. Teacher attitudes include satisfaction with and preparedness for the job as well as perceptions of school climate, including the collaboration and competition among staff. We are interested in teacher experiences with respect to evaluation, professional development intensity, mentoring support, and accountability. These impacts on intermediate outcomes can help bridge the divide between our implementation analysis and our analysis of impacts on final outcomes.

Experimental Design and Analysis

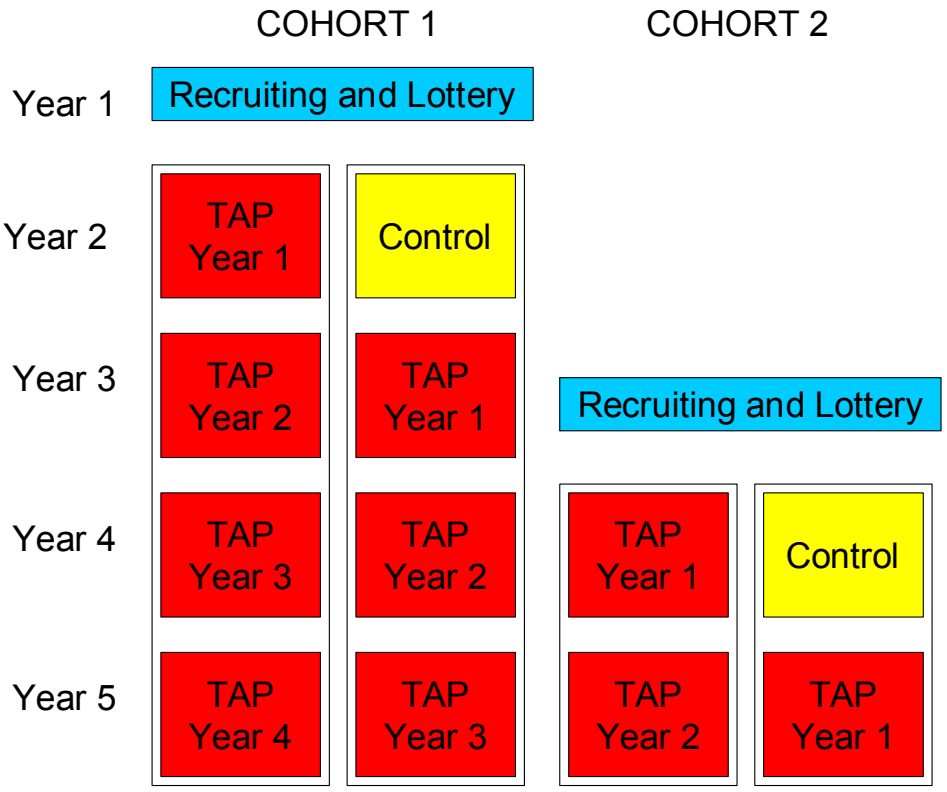
The experimental design will provide unbiased estimates to address all the impact questions as well as the third implementation question regarding differences in school practices. The experimental component of the evaluation is illustrated in simplified form in Figure 1. During the planning year (Year 1), CPS recruited 20 schools to participate in TAP, and MPR assigned half of the schools to a treatment group and half to a control group. As discussed below, most assignments were made at random using a lottery. The treatment group will begin implementing TAP in 2007 and the control group will delay TAP implementation until 2008. Because the Chicago Public Schools required at least one high school and at least one charter school per cohort, and was unable to allow random assignment of high school and charter school, we randomly assigned eight schools to each cohort. Thus each box in Figure 1 can be thought to represent the eight elementary (kindergarten through grade eight) schools.[3]

During program Year 3 (2008 to 2009), CPS will recruit another cohort of 20 schools, thereby doubling the sample size for estimating impact of the first and second years of TAP implementation. MPR will randomly assign those schools to treatment (TAP) or control (delayed implementation of TAP), and MPR will follow these additional cohorts for two years while continuing to follow the first cohort for two additional years.[4]

[3] We randomized 8 schools per group, but will include 10 schools per group in the data collection. Each group of 10 schools includes one charter school and one high school, but CPS was not willing to allow these schools to be randomly assigned to implementation year.

[4] If more than 20 schools volunteer for the second lottery, we will use random assignment to create a third group of schools that will not implement TAP in Year 4 or 5. This third group will receive priority for future implementation after the study ends.

Figure II.1. Experimental Design of an Evaluation of TAP in Chicago



An important consideration in the interpretation of findings from our study will be startup effects discussed in Chapter I. The TAP model is well established, having been in schools since 2001. However, TAP was introduced to the Chicago Public Schools in 2007, and rules for the exact implementation were finalized as late as summer 2007, just prior to implementation in the fall. For this reason, the impact estimates we generate will contain a startup effect. A small or negative startup effect does not necessarily imply that the steady state impact will be small or negative, but would provide cautionary evidence on the short-term effects of introducing such a program.

In the 2007 lottery, we used simple random assignment with unequal selection probabilities, based on school readiness ratings given by CPS. That is, CPS determined that the candidate schools varied in their readiness to implement the program in 2007. Prior to the lottery, CPS rated all of the candidate schools on a three-point scale: (1) definitely ready, (2) probably ready, or (3) possibly not ready. We selected from the first group with a probability of 3/4, from the second group with probability 4/7, and from the third group with probability 1/5. Thus, we can weight the schools by the inverse selection probability to compute the estimated impact of TAP on a school of "typical" readiness and use the variance of these weights to calculate the reduction in precision that results from this approach.

In 2009, we will take advantage of what we expect will be a larger pool of schools that are "ready" to implement and more lead time once the eligible schools are identified. With

these advantages, we will use a method of random assignment known as block stratified randomization. That is, we list a set of characteristics believed to be related to outcomes, such as average teacher experience and prior student test scores, and we create pairs of schools that are as similar as possible. Then we select one member of each pair at random to be in the first cohort and one member to be in the second cohort. The stratified randomization increases the statistical efficiency of the estimates.

Threats to the Integrity of Random Assignment Design

It is important to acknowledge and understand the limitations of the experimental design. Most important, the sample size is limited. We expect to have eight elementary schools per cohort for the experimental analysis. The concern with small numbers of schools is statistical power, or our ability to detect an impact given that one exists. Similarly, differences between TAP and non-TAP schools may include both TAP impacts, which we wish to estimate, and other confounding factors that differ by chance. For example, if the eight TAP schools happen to have more talented principals than the eight non-TAP schools, then the estimated impact would be biased upward. To address this concern, we will describe the schools—including students, teachers, and administrative staff—in great detail and document any differences that might be confounded with treatment status. We will control for easily observed and objectively measured factors using regression analysis (presented in Chapter IV) and will use more subjective or qualitative information to provide additional interpretation of the findings.

Another potential concern is contamination that occurs when the presence of a treatment affects the control condition. The main assumption of the random assignment design is that the control group represents what would have happened in treatment schools if treatment did not exist at all. If behavior in the control is affected by the treatment (either the treatment schools or the control schools' prospect of implementing the treatment in the future) in any way that also affects outcomes, then the assumption is violated. For example, if TAP Mentor or Master Teachers are hired away from control group schools, then the control schools would be worse off as a result of the treatment and the resulting impact estimates would be upward biased. Similarly, the presence of TAP schools in the same district might induce control school teachers to attempt to transfer into the treatment schools or it might influence a prospective teacher to choose the TAP school instead of the control school where she might have gone instead in the absence of TAP. Another possibility is that control schools may recruit better teachers than they would have otherwise, because they can offer prospective teachers the chance to participate in TAP in their second year.

In these scenarios, the control group is affected by the presence of treatment and hence does not accurately represent the counterfactual. To address this concern, we will collect lists of mentor/master teacher candidates interviewed by TAP schools and determine whether any control school teachers applied. We will track teacher and principal mobility prior to and after the random assignment announcement. We will include a question on the teacher survey asking whether teachers who are new to the TAP school had applied to, interviewed for, or considered a position in any other schools to determine if they might

have gone to a control school. Finally, we will interview control school principals (in addition to treatment school principals) to discuss the role that TAP or the future promise of TAP might have had in influencing hiring. By carefully documenting all the possible sources of contamination we can speculate on the size and direction of any possible bias and use that information to interpret the impact estimates.

One final concern is the foreknowledge of treatment assignment on behalf of the control schools. If control school principals and teachers know they are slated to implement TAP in the following year, they may forego alternative reforms that they might otherwise have implemented, choosing instead to wait until they implement TAP. Alternatively, having learned about the program in order to decide to adopt it, the control school may attempt to implement components of TAP before the program is formally introduced. This anticipatory behavior is not likely to threaten the validity of the study because the program is fairly resource-intensive, and thus it would be difficult to implement the central tenets of the program prematurely. In terms of forgoing alternative reforms, we will explore whether schools that were not selected to enter the lottery but expressed interest have adopted any programs that have elements in common with TAP.

Quasi-Experimental Design and Analysis

MPR will supplement the experimental analysis with a quasi-experimental analysis, thus making the most of administrative data on test scores and teacher mobility for the approximately 380 CPS elementary schools that are not participating in the experimental study. The quasi-experimental analysis amounts to constructing a synthetic control group by matching non-TAP schools that did not volunteer for the study to TAP schools based on prior growth in student achievement, teacher background, percent low-income, and several other school variables. In particular, MPR will identify matches by using a propensity score, which is the predicted probability from a statistical model of being a TAP school based on these characteristics. Because this approach relies on an assumption that we know and can observe the determinants of teacher and student outcomes, it does not provide the same protection against selection bias as a randomized experiment. Nevertheless, it uses a much larger sample and allows the evaluation to address a wider range of questions than is possible with the experimental approach alone. For example, propensity score matching can be used to estimate TAP impacts in schools that implemented the program with greater or lesser fidelity to the model.

We will use the quasi-experimental methods to supplement the experimental estimates, addressing the same impact questions with information from a much larger sample. While the quasi-experimental methods require reliance on assumptions not needed for the randomized experiment, they provide independent information from a large sample, and thus serve as an important check on the main findings. In particular, the quasi-experimental sample will be useful for estimating impacts for subgroups (impact research question number 4).

Challenges and Limitations of the Quasi-Experimental Design

As with the experimental comparisons, the quasi-experimental comparisons have some important limitations to keep in mind. First, many of the same concerns about the randomized control group are also relevant for the matched comparison group. For example, contamination is still a concern, even if it is to a lesser degree. Mentor and Master teachers might have been hired from the comparison schools. Classroom teachers or other building staff might be induced to leave a comparison school to attend a TAP school if they find that model attractive, and the movers could be more talented or productive than non-movers. As with the experimental analysis, we can consider this mobility to be part of the treatment and conduct the analysis as we would otherwise. We can observe the mobility using district data on teacher assignments by year and thereby infer the degree to which any observed impacts are the result of teachers improving their practice (productivity effects) or more talented teachers seeking to work in TAP schools (composition effects).

A concern that is unique to the quasi-experimental analysis is the possibility of unmeasured characteristics of TAP schools that we cannot capture in our model. As mentioned above, we will conduct qualitative research to provide some information on the possibility that treatment effects are the result of selection rather than true school improvement. This will include learning from Project REAL and NIET staff how schools were selected and recruited in order to understand the process of selection into the study to inform the matching process. We will conduct sensitivity analyses to test whether variations in the implementation of the matching method might make a difference in the study's findings. We also have the benefit of the randomized control group, which we can compare directly to the nonrandomized matched comparison group. If the outcomes for the control and comparison groups appear similar, it will lend credence to both designs. If the outcomes are divergent, we will use statistical modeling and qualitative evidence to reconcile the differences.

IMPLEMENTATION AND FIDELITY ANALYSIS

The implementation analysis will focus on TAP schools, but will include comparisons with non-TAP schools in order to contrast TAP implementation with the related teacher development practices normally found in schools. We also wish to compare TAP schools in Chicago with TAP schools nationwide as a second point of reference to determine whether Chicago's TAP experience is typical.

We are working with NIET staff to have them share with MPR the data they will collect on program implementation. These TAP-specific measures will be used to assess the degree to which the program was implemented in schools that were assigned to implement it. Other measures will be more generic, covering elements such as teacher evaluation, and professional development, and teacher-leader roles (such as mentoring) that would be present in TAP schools but might also be present to varying degrees in non-TAP schools. We will also gather data on teacher evaluation, compensation, and working conditions.

We will collect these data through teacher surveys and principal interviews in project Years 2 and 4 for all schools that entered the lottery, including TAP and control (delayed TAP) schools, as well as a subset of non-TAP comparison schools that did not participate in the lottery. As we discuss in Chapter III, we decided on Years 2 and 4 because it allows us to capture the program experiences after the first year of implementation (for treatment groups) and include a non-implementing control group. We collect the data at the end of the year to be able to capture program experiences retrospectively and attitudes after some time has passed. We would prefer to conduct additional years of surveys but the resources for primary data collection are limited.

To complement the surveys, we intend to review program documents and interview program staff on a regular basis to help us understand implementation as it happens. This will include meetings with the Project Manager of REAL, the NIET staff working with Chicago, as well as CPS officials who work in other capacities that relate to teacher hiring, evaluation, and support.

Limitations of the Implementation Analysis

It is important to understand some limitations of the implementation analysis. Principally, it will not be possible to document every aspect of program implementation. Many of the program experiences may be unique to Project REAL, unique to the Chicago setting, or both. The project will not include formal observations of TAP because such observations are not feasible within the study's budget constraint.

Another important limitation is that any statements we make about implementation "quality" or barriers/facilitators of good implementation will require some subjective interpretation. Our primary informants will be CPS teachers and principals, but we will also rely on a review of program documents and a set of interviews with program staff within CPS and NIET. Starting up a highly complex program anew will have many challenges that are unique to the setting in which the program is being implemented, so we will focus on both issues that can help program staff improve the program for future implementation as well as general lessons for programs around the country.

CHAPTER III

DATA COLLECTION

To address the study's research questions, we will obtain several types of information from a variety of sources. The sources of primary data (i.e., data collected by Mathematica specifically for the evaluation) will be a teacher survey and interviews with principals and program staff. The evaluation will also make extensive use of existing data sources, including CPS administrative data on students' scores on state assessments, student backgrounds, and teacher mobility and credentials, as well as NIET data on the implementation of the TAP program gathered from classroom observations, program reviews, and a teacher attitude survey.

Figure III.1 displays a timeline for the data collection activities. The matrix presented in Table III.1 maps the data collection into the research questions being addressed. "HLM" refers to hierarchical linear model explained in Chapter IV. A description of each activity is provided below. Draft instruments can be found in the appendices. Prior to any data collection, MPR will prepare and submit a request for approval to CPS's Institutional Review Board (IRB) for Research Involving Human Subjects if applicable. The submission will include descriptions of research procedures, copies of data collection instruments and protocols, and estimates of the burden associated with each.

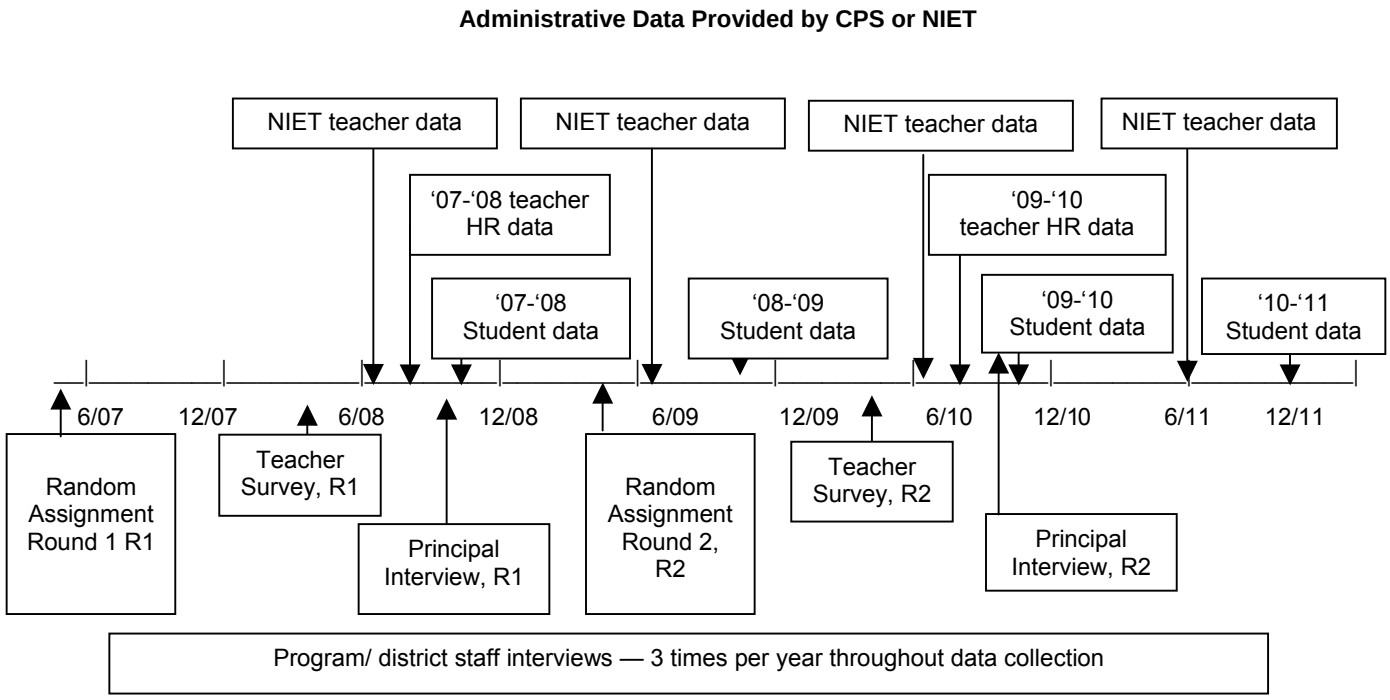
TEACHER SURVEY

Rationale

Teachers are key informants on both the impact and implementation research questions because they directly experience the elements that characterize TAP school reform, and are first-hand observers of the effect TAP is having on the school staff and overall community. The teacher survey will gather detailed information about teachers' professional background and experience that is not available in the CPS administrative records (see Appendix A for the draft survey). It includes six sections focusing on (1) teachers' educational background and professional experience, their certification status, and their current teaching assignment; (2) the types of professional development and support that teachers receive at their schools; (3) the leadership roles and responsibilities teachers have assumed in addition to their regular classroom teaching duties; (4) the compensation, or potential for compensation, that teachers receive for their performance and that of their students; (5) teachers' attitudes and

satisfaction with various aspects of their school and the opportunities provided to them; and (6) teachers' basic demographic characteristics.

Figure III.1. Data Collection Timeline



Sampling

The survey will be administered to teachers in the 10 schools that are implementing the TAP program in 2007-2008, the 10 control schools assigned to delay implementation of the program until the 2008-2009 school year, and a subset of 20 matched comparison schools that will not implement TAP.[5] We will randomly sample 20 teachers from each school for a total of 800 teachers. Table III.2 illustrates our proposed sample allocation of teachers by specific characteristics. Our goal is to include teachers in the sample whose data we can match to the student test scores provided by CPS.[6] These teachers will be referred to as Round 1 teachers. We will repeat this data collection with teachers in the 10 schools

[5] Only 16 of the 20 schools assigned to treatment status were assigned at random. We will gather data on the 16 randomly assigned schools as well as the 4 purposively assigned schools (two high schools and two charter schools) even though we can only estimate experimental impacts using the randomly assigned schools.

[6] Teachers will be assured that their responses will be kept strictly confidential, will not be shared with CPS, and will only be reported in aggregate with those of other respondents.

Table III.1. Research Questions, Data Items, and Sources

Research Questions	Data Items	Data Source	Data Analysis
Impact of TAP on teachers and students			
1. What is the impact of TAP on teachers' "value added" to student achievement and on career decisions (transfers and retention)?	Student achievement in reading, math, and science Teacher mobility patterns, reasons for school departure, satisfaction with current school	Scores on the ISAT Teacher Survey Principal Interview CPS data on teacher mobility	HLM, Difference in mean outcomes
2. How much of the TAP impact on teacher quality is a productivity effect (teachers doing their job better) versus a composition effect (improving the mix of teachers)?	Student achievement in reading, math, and science Teacher mobility and retention	Scores on the ISAT Principal Interview CPS data on teacher mobility	HLM estimated separately for full sample and stayers
3. Does the program become more effective over time as the school has more experience adapting to it?	Students: Changes in achievement over time Teachers: Changes in teacher mobility and satisfaction over time	Scores on the ISAT Teacher Survey Principal Interview	HLM with time x treatment interactions
4. Does the program have greater impacts for some types of schools or teachers than others?	Schools: Location, proportion of low-income students, proportion of LEP students, average student achievement scores in reading and math; other policy initiatives or governance issues; principal characteristics Teachers: Teacher's education, qualifications, years teaching, professional development	CPS student data CPS teacher background data Teacher Survey Principal Interview Program Staff Interviews	HLM with subgroup analysis
Implementation of TAP			
1. Was TAP implemented faithfully according to the model outlined by its developers?	TAP-specific characteristics (professional development and support; performance-based compensation, attitudes toward TAP)	NIET Teacher Survey NIET Site Visit Reports Principal Interview Program Staff Interviews	Comparison to program model and non-CPS TAP schools
2. What factors facilitate or impede the successful implementation of TAP?	School-level barriers and supports to TAP implementation	NIET Site Visit Reports Program Staff Interviews	Qualitative analysis of interview data
3. How does TAP differ from what is normally implemented in CPS schools?	Teachers' professional development and experiences with mentor/master/coach teachers; support received from school administrators and other teachers	Teacher Survey and Principal Interview in non-TAP schools	Difference in mean outcomes; Analysis of qualitative data

Table III.2. Proposed Allocation of Teacher Sample

School level	Number of Schools/Teachers			In-Sample Teacher Characteristic	Out-of-Sample Teacher Characteristics
	TAP	Control	Com-parison		
K-8 (36)	9/180	9/180	18/360	School-based teachers of academic subjects Teach grade level in which students are administered state assessment (80%) Provide academic support (special education and resource teachers) (20%)	Teachers of non-academic subjects
High School (4)	1/20	1/20	2/40	School-based teachers of math, English, or science or history (80%) Mentor teachers (20%)	Teachers of elective subjects Special education teachers
Total (40)	10/200	10/200	20/400		

assigned to implement the TAP program in Year 4, along with those in the schools assigned to the control group in that year and in a set of 20 comparison schools. These will be referred to as Round 2 teachers.

Procedures

Timing. The teacher survey will be conducted in the spring of 2008 and 2010, the first year of TAP implementation for treatment schools in each round. The surveys will be mailed out in late March and telephone followup with nonrespondents will take place beginning in late April and continuing through May. Ideally, we would administer a survey every year. However, given the study's limited resources, we chose the most critical time point, which falls right at the end of the first implementation year, before the control schools have begun implementing TAP. While it would be helpful to administer a survey to teachers at baseline—before any school implements TAP—in order to track changes in school-level variables over time, this is not strictly necessary given the randomized control group. The control group already tells us what would have happened in treatment schools in the absence of treatment.

Mode. The mode of data collection will be a mail survey with telephone follow up. In our experience with large and small teacher surveys, we have found this methodology to be the most cost-effective way to produce high response rates (over 80 percent). We considered other modes—for example, conducting in-person interviews to collect data from teachers—but we recommend against this approach for two reasons. First, the types of information that we are requesting are generally closed-ended items that can be most efficiently answered in a self-administered written questionnaire. Second, teachers' work schedules make it difficult to contact them to schedule appointments and find time when they are free to participate in an interview. Inviting them to a single large session for group

administration would be impractical because there is no convenient reason to convene the teachers that is *not* related to treatment status. Even with group administration, some follow up would be necessary for absent or non-compliant teachers, and the potential for nonresponse bias (resulting from a decidedly non-random subset of teachers failing to complete the survey) is significant. Finally, we do not recommend a web-based survey because we have found that for a sample of this size, a web-based survey is not cost effective given that it only yields about a 30 percent response rate, with the remaining data collected through mail and phone follow up. Therefore, we concluded that a mail survey with follow-up phone calls was most suitable.

In March, we will mail the survey to teachers at their schools, along with a letter explaining the purpose of the survey and assuring confidentiality. Teachers will be asked to complete it within two weeks and return it in the pre-addressed, postage-paid envelope included in the survey packet. We will conduct telephone follow up for nonresponse and data retrieval. The survey will take about 30 minutes to complete, and teachers will receive a $25 incentive—for example, a gift card at a chain bookstore—when their completed survey is returned. We anticipate these procedures will yield a final response rate of 80 percent (about 35 percent of those by mail, and 65 percent through telephone follow up).

PRINCIPAL INTERVIEW

Rationale

The role of school principals in the implementation of a reform such as TAP is significant, and their assessments of the impact are key to evaluating TAP's success in Chicago. As school leaders, principals are ongoing witnesses to their school's climate, teachers' job satisfaction and motivation to improve and advance their careers, and the impact that a whole-school reform program may have on these factors. More relevant to the TAP reform is the role that the school principal has in evaluating teacher performance and recommending performance-based compensation. For these reasons, this study must collect data from school principals.

TAP principals are asked annually by NIET to complete a survey assessing how TAP has changed their school and how well the different elements of the program are implemented in their building. In addition, principals report on teacher turnover rates in their schools and the main reasons for teacher departure. To complement these NIET data and CPS data on teacher mobility, MPR will interview the principals of the 10 TAP and 10 control schools. The interview will collect information on the following topics: school context, teacher hiring and assignment practices, teacher evaluation and supervision policies, degree of teacher mobility at the school, and the principal's professional background (see Table III.3 for a draft list of interview topics).

Procedures

Timing and Sample. The principal interviews will be conducted at the start of the second year of TAP implementation, in the second and third weeks of October. This time was selected because principals in the TAP schools will have had a year of program implementation and will be able to provide a retrospective report on their school's transition to a TAP school. We will therefore interview Round 1 principals in fall 2008 and Round 2 principals in fall 2010.

Table III.3. Principal interview topics

I. School Context

- Organization of classes and teacher collaboration
- Structure of the school day and release time for mentoring
- Special policy initiatives (including TAP)
  - Follow up: How did the school come to participate?
  - Follow up: Implementation experiences? (probe in detail)
- Governance issues
- Teacher collegiality
- Changes in school climate over prior year
- Crises that occurred during the year

II. Teacher Hiring and Assignment

- Number of vacancies by assignment
- How/when filled
- Hiring process, factors that influenced hiring success
- Assignment to grades/subjects
- Filling leadership positions in the school
- Unexpected challenges in hiring and assignment

III. Teacher Evaluation and Supervision

- Beginning teacher supports
- Professional development
- Curriculum support staff
- Release time and planning time, common planning time
- Direct observation of classroom practice and feedback to teachers
- Other evaluation and feedback
- Unexpected challenges in program implementation, teacher evaluation, and supervision

IV. Teacher Mobility

- Teachers who did not return for 2008-2009 (2010-2011)
- Circumstances of their leaving, where did they go, when did the principal find out
- Circumstances of filling vacancies

V. Principal Background

- Years as an administrator (this school, total)
- Years as a classroom teacher (grade/subject assignment in which served the most years)

Mode. The interview will be conducted by telephone and will take approximately 45 minutes to complete. Because the questions for principals will be primarily open-ended, an interview is the most efficient means of collecting these data, and also allows for follow up for purposes of clarification and elaboration, if necessary. We recommend telephone interviews rather than in-person interviews because of the nature of principals' work days, where it is not uncommon for unexpected events to necessitate rearranging their schedules. It is much easier to reschedule an interview, or complete an interrupted interview, when it is conducted by phone.

MPR has expert interviewers on staff with experience interviewing education officials. We will provide rigorous study-specific training to ensure that interviewers are knowledgeable about CPS and TAP, can explain the importance of the survey to the sample member, answer any questions, and gain cooperation from those who initially refuse to participate. Responding principals will also receive a $25 incentive for their participation.

ADMINISTRATIVE DATA FROM CPS

We will measure student achievement, the primary outcome for this evaluation, using student assessment data collected routinely by CPS. These data include students' scores on state assessments in mathematics and reading for grades three through eight and in science for grades four and seven, plus any scores routinely available for high school students. We will also use student background information, such as race, gender, free/reduced-price lunch eligibility, enrollment status (to track their movement between schools and in and out of CPS), and disability or special education status. We will collect these data for each year of the study, including the year prior to the first lottery (2006-2007). We have assumed that the data will link every student by unique ID code across years to a school and grade level. Our analysis will be richer if we can also link students to the teachers responsible for each test subject; however, we can still proceed if such linkages are not in place for all schools in all years.

Information on teachers' credentials, years of experience, and teaching assignment (to track mobility) will also be gathered from CPS human resources records. We will request any other teacher-level variables that might be correlated with teacher performance or mobility. We will request these data for all CPS teachers during 2006-2007 through 2010-2011.

TAP IMPLEMENTATION DATA FROM NIET

To assess TAP implementation, we will use the implementation data collected annually by NIET for all TAP schools in Chicago. These include findings from classroom observations, review of program data, results of a teacher attitude survey, and findings from a principal interview. The data may be provided in raw or aggregate form, based on discussions with NIET. The goal of this data collection will be to obtain scores that summarize different dimensions of fidelity to the TAP model as well as teacher attitudes about TAP, so we can relate these to other sources of data we collect for the evaluation.

INTERVIEWS WITH NIET AND CPS STAFF

Three times per year (from fall 2007 through fall 2011), MPR will conduct semi-structured interviews with the NIET and CPS staff responsible for implementing and monitoring TAP in Chicago. The topics of these interviews will be determined by the study activities at the time of the interview. Examples of interview topics include: recruitment of schools for participation in the study; training on TAP implementation and any issues that arise during or as a result of the training; unintended or unexpected consequences linked to the results of the random assignment or schools' participation in the study; and informal observations on the impact of TAP on teachers, administrators, or other stakeholders in CPS.

CHAPTER IV

ANALYSIS AND REPORTING

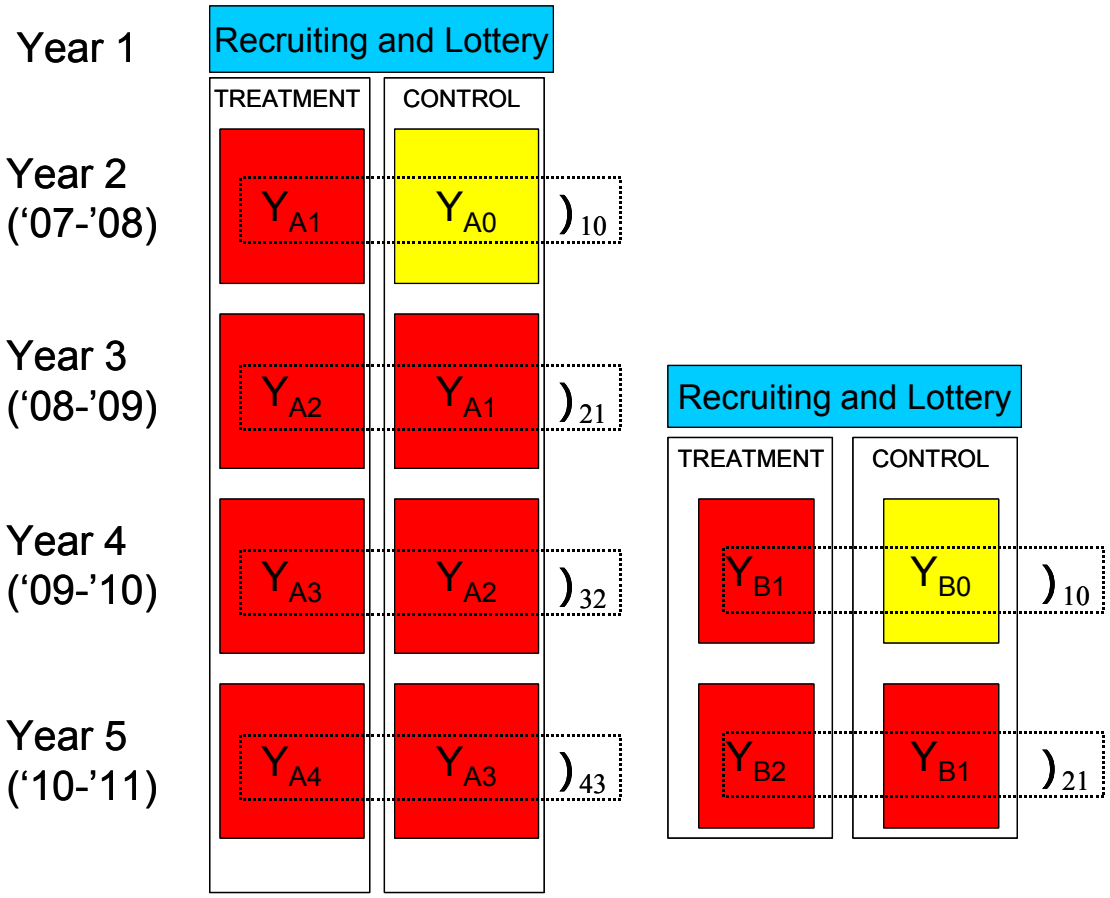
The main goal of the data analysis will be to estimate the impacts of TAP on student achievement and the recruitment and retention of teachers. We will also study the implementation of TAP in order to obtain information critical for interpreting the impact estimates and to draw lessons useful to stakeholders who are considering implementing or expanding TAP in the future. This chapter discusses the experimental and quasi-experimental methods that we will use to estimate impacts and our approach to the implementation analysis. We also discuss the reporting and dissemination of the findings.

EXPERIMENTAL IMPACT ANALYSIS

An important feature of the study design is MPR's assignment of schools by lottery to either a treatment group that begins implementing TAP in the fall following the lottery or a control group that delays implementation by one year (see Chapter II). With this experimental (randomized) design, the differences between average outcomes in the treatment and control schools yield unbiased estimates of the impacts of TAP. Specifically, at each year following the random assignment lottery, the difference in means provides an unbiased estimate of the impact of one additional year of TAP. Random assignment ensures that there are no *systematic* differences between treatment and control schools prior to starting TAP; we use statistical hypothesis tests (for example, a t-test for the difference in two means) to account for *chance* differences between treatment and control schools. Figure IV.1 illustrates the experimental comparisons that can be made each year following the lottery. The interpretation of these comparisons is discussed below.

Study Year 1 is for planning and recruiting. At the end of Year 2—when treatment schools have experienced one year of TAP and control schools have not implemented TAP—the treatment-control difference, denoted Δ_{10}, is an unbiased estimate of TAP's impact in its first year. That is, the difference between the mean outcome for treatment schools (in Figure IV.1, the top red box representing eight schools) and the mean outcome for control schools (the adjacent yellow box, representing the other eight schools assigned in 2007) estimates the extent to which one year of TAP implementation improves the outcome relative to what would have occurred in the absence of TAP. We note that we will ultimately be able to double the number of schools used to calculate this impact: 16 schools randomly assigned in the 2007 lottery and another 16 schools randomly assigned in the 2009 lottery.

Figure IV.1. Experimental Comparisons



Note: One red or yellow box represents approximately eight schools.

Comparisons of treatment and control mean outcomes in subsequent years yield unbiased estimates of the *incremental* impact of each additional year of TAP. For example, at the end of study Year 3, treatment schools will have implemented TAP for two years while control schools will have implemented TAP for only one year; thus, the difference in the experimental groups, Δ_{21}, estimates the impact of a second year of TAP implementation relative to a single year. These incremental impacts provide evidence on whether TAP becomes more effective over time as schools gain experience adapting to the program. If the full impact of TAP occurs after only one year of implementation, then we would expect zero differences, on average, between treatment and control schools after the latter undergo a year of TAP implementation; that is, $\Delta_{21} = \Delta_{32} = \Delta_{43} = 0$. Alternatively, positive differences between treatment and control schools in these later follow-up periods would suggest that the benefits of TAP continue to accrue with additional years of implementation (a program maturation effect discussed in Chapter I), while negative differences would suggest that the benefit diminished over time. For example, a positive Δ_{21} would indicate an improvement in average outcomes during the second year of TAP implementation beyond

the improvement experienced during the first year of implementation.[7] The sum of successive impact estimates, for example $\Delta_{10} + \Delta_{21}$, represents our best experimental estimate of the effect of a program in its second year relative to no program at all.

Statistical Model

Building upon the basic differences-in-means approach, we will use a hierarchical linear model (HLM) to compute regression-adjusted estimates of the impacts of TAP. The HLM accounts for the nesting of students within classrooms and classrooms within schools. Using regression procedures also increases the statistical precision of the estimates by enabling us to account for student, teacher, and school characteristics other than TAP status that could affect the outcome.

We will use a basic three-level HLM that consists of student, classroom, and school components:

(1) Student level: $$Y_{ijk} = \alpha_{0jk} + \alpha_1 X_{ijk} + \varepsilon_{ijk}$$

(2) Classroom level: $$\alpha_{0jk} = \beta_{0k} + \beta_1 W_{jk} + \mu_{jk}$$

(3) School level: $$\beta_{0k} = \gamma_0 + \gamma_1 T_k + \gamma_2 Z_k + \nu_k$$

where Y is the outcome of interest (for example, student test score) for student i in classroom j in school-grade k; X is a vector of time-invariant student characteristics or time-varying characteristics measured at baseline, which we call "exogenous"; W is a vector of exogenous teacher or classroom characteristics; T is a treatment status indicator that equals 1 if the school was assigned to the treatment group and equals 0 otherwise; Z is a vector of exogenous school characteristics; and, ε, μ, and ν are random (and mean 0) student-level, classroom-level, and school-grade-level errors, respectively.

The three levels can be combined in a single model as follows:

(4) Combined model: $$Y_{ijk} = \gamma_0 + \gamma_1 T_k + \gamma_2 Z_k + \beta_1 W_{jk} + \alpha_1 X_{ijk} + \psi_{ijk}$$

where the composite error term, ψ, is defined as $\psi_{ijk} = \nu_k + \mu_{jk} + \varepsilon_{ijk}$.

Equation (4) shows that the individual student test score (for example) is modeled as a function of the average achievement of students in the control schools (γ_0); the impact of TAP for students in treatment schools (γ_1); school, teacher/classroom, and student

[7] As with Δ_{10}, we can increase the sample size for estimating Δ_{21} and, consequently, the precision with which we estimate this impact by using both the schools assigned in the 2007 lottery and the schools assigned in the 2009 lottery.

characteristics; and an error term with school, classroom, and student components. We presented a three-level model with student-level outcomes, but this framework can easily be simplified to a two-level model for outcomes measured at the teacher or classroom level, such as teacher retention, or as a one-level model for school outcomes. We continue to discuss the three-level (student) model below as an example.

The estimate for γ_1 represents the regression-adjusted estimate of the impact of TAP, the main parameter of interest. As discussed above, the interpretation of this impact depends on the year for which the impact is estimated. For example, if estimated at the end of study Year 2 (summer 2008) when the treatment schools have had one year of TAP and the control schools have had no TAP, the estimate for γ_1 represents the impact of one year of TAP relative to no TAP. If estimated at the end of study Year 3 (summer 2009) when treatment schools have had two years of TAP and control schools have had one year, the estimated coefficient indicates the incremental impact of a second year of TAP relative to the first year.

Our three-level model assumes that we will have individual-level student data linked to teachers. We anticipate being able to obtain such data for treatment schools because linked data are necessary to determine the performance-based compensation given under TAP. We assume we will also obtain student data matched to teachers for control schools and for at least an identified subset of non-TAP Chicago public schools serving as comparison schools. Failure to link students to teachers will hinder our ability to account for teacher and classroom characteristics, resulting in a slight loss of statistical precision, but we will still be able to link students to their school and grade. In that case, we would be able to consider teacher and classroom controls, but only at their average levels for the school/grade combination.

The estimates of the impacts on student achievement are necessarily restricted to tested grades and subjects, which we assume will be math and reading in grades 3 through 8. While we propose to estimate the model separately by grade level, we will also aggregate the grade-specific impacts to generate an overall effect. While some testing is done at the high school level, we will not be able to generate impact estimates for those grades. The main reason is that CPS did not want to randomly assign high schools. Even if we were to have randomly assigned the high schools, or conducted nonexperimental analysis of high school data, there is only one high school per cohort, which will not support statistically generalizable findings. Nevertheless, if CPS provides high school-level data, we will report mean student achievement outcomes for TAP and non-TAP high schools with the proper caveats that such differences reflect differences in the schools as well as differences caused by the TAP program itself and we cannot necessarily separate those two types of influences.

Although we illustrate a model that allows for student, teacher/classroom, and school characteristics, the relatively small number of schools and classrooms in the sample limits the number of such characteristics that we can include as explanatory variables. Among the most important baseline variables to include will be baseline measures of the outcome—for example, when estimating the impact of TAP on student achievement we will include prior

year test scores. Such variables are likely to be highly correlated with the outcome measures and, consequently, will improve the precision of the impact estimate.

The statistical method we used to estimate equation (4) on page 25 will depend on the nature of the outcome variables. The overall estimation follows the set of procedures available for all mixed models, which contain both fixed and random coefficients. We will use restricted maximum likelihood (REML) procedures for the general linear mixed model for continuous outcome measures (such as student test scores) and maximum likelihood logit methods for binary outcome measures (such as an indicator for whether a teacher remained in the school the next year).

Longitudinal Analysis

In addition to estimating a separate impact for each period, we can expand the model to conduct a longitudinal analysis and test hypotheses about the pattern of incremental impacts over time. At the end of Year 5 of the study we will be able to estimate a model that includes four years of impact estimates. Specifically, we can stack observations from all available time points and estimate the following combined model:

$$(5) \qquad Y = \gamma_0 + \gamma_1 T + \sum_{t=2}^{4} \delta_t F_t + \sum_{t=2}^{4} \lambda_t F_t \times T + \gamma_2 Z + \beta_1 W + \alpha_1 X + \psi$$

where F_t is a follow-up period indicator that equals 1 for observations pertaining to follow-up period t after random assignment and 0 otherwise, $F_t \times T$ denotes the interaction between the follow-up period indicator and the treatment status indicator, and all other variables are as defined previously. We have omitted subscripts for simplicity of presentation.

In this longitudinal model, the estimate for γ_1 now represents the estimated impact of TAP in the first follow-up period, which is the 2007-2008 school year. The estimate for λ_t represents the difference between the impact at time t and the impact in the first follow-up period. For example, the estimate for λ_3 represents the difference between the impact in the third follow-up period (the 2009-2010 school year) and the first follow-up period. Statistical tests involving linear combinations of the estimated coefficients can be performed to test hypotheses such as whether the impact in the third follow-up period (estimated by $\gamma_1 + \lambda_3$) differs from the impact in the second follow-up period (estimated by $\gamma_1 + \lambda_2$). Combining coefficients will also enable us to explore the accumulation of impacts over time—for example, adding the incremental impact of the second year of TAP to the first-year impact will provide an estimate of the full impact of two years of TAP.

We will consider additional models in which the TAP effect grows linearly, quadratically, or according to some other curvilinear relationship with time. Equation(6) shows a quadratic form.

$$(6) \qquad Y = \gamma_0 + \gamma_1 T + \delta_1 F + \lambda_1 T \times F + \lambda_2 T \times F^2 + \gamma_2 Z + \beta_1 W + \alpha_1 X + \psi$$

Using the model in equation (6), we can test whether outcomes grow or accelerate at a faster rate with exposure to TAP.

QUASI-EXPERIMENTAL ANALYSIS

As discussed in Chapter II, we will supplement the experimental analysis with a quasi-experimental analysis. We plan to use a type of matching to identify non-TAP CPS elementary schools that are similar on observable characteristics to the TAP schools. The matching is based on a "propensity score," defined here as the predicted probability of being a TAP school, based on a specified set of observable characteristics. To implement this method we will first estimate a statistical model of TAP participation and then match schools that are nearby in a statistical sense based on the propensity score to identify the comparison group. By instituting different rules for estimating the propensity score and for defining "nearby," we can test the method's robustness to different assumptions we make in implementing it.

To generate the propensity scores, we will estimate a logit model to predict schools' participation in TAP using school characteristics. The dependent variable is an indicator for participating in TAP, coded as 1 for schools that participate and 0 for schools that do not. The explanatory variables will include prior growth in student achievement, teacher background, percent low-income, and other school variables. We will use the estimated coefficients to generate a propensity score for each participating and nonparticipating school. With help from Project REAL staff, we can document the process by which TAP schools were selected and thereby improve our ability to model the selection/participation process.

Once propensity scores have been estimated, several matching algorithms can be used to identify appropriate comparison schools for the TAP schools. For example, nearest neighbor matching involves choosing for each TAP school the non-TAP school with the closest propensity score, which is its "nearest neighbor" in a statistical sense. It is possible, however, that the propensity score of the nearest neighbor differs substantially from that of the TAP school. Variants of nearest neighbor matching reduce this possibility by allowing a TAP school to be matched only to a non-TAP school (or set of non-TAP schools) whose propensity score falls within a pre-specified distance of the TAP school's score. An alternative matching strategy called kernel matching would make use of all non-TAP schools in forming a comparison group. Under kernel matching, each TAP school would be matched to a weighted average of all the non-TAP schools, with weights depending on the distance between propensity scores of the TAP and non-TAP schools.

We note that the practical consideration of the ease with which we can obtain data for non-TAP schools will affect the matching strategies we can use. Using nearest neighbor matching or one of its variants would enable us to identify a subset of non-TAP schools for which we would need to collect follow-up data. Kernel matching, however, would require obtaining data for a larger number of non-TAP schools and is therefore most practical when all of the variables in the analysis are generated from administrative records.

Employing different matching algorithms will likely produce different comparison groups. The quantity and quality of matches can vary across strategies, often resulting in tradeoffs between bias and statistical precision. To help assess the robustness of our estimates, we intend to explore several matching strategies and present results for multiple comparison groups, reconciling any differences should they emerge.

Quasi-experimental estimates of impacts can be obtained using the same equations described for the experimental analysis. Rather than indicating the experimental condition to which a school belongs, the treatment indicator *T* in the quasi-experimental context indicates whether a school belongs to the treatment group or the comparison group identified via propensity score matching. The difference in mean outcomes between treatment and non-experimental comparison schools estimates the impact of *t* years of TAP relative to no TAP, but these quasi-experimental estimates may be biased by unobserved differences between the TAP and non-TAP schools.

Hybrid Model

We can combine the experimental and quasi-experimental analyses into a single model to facilitate the statistical testing of relationships of interest among impacts. This hybrid model is estimated using observations from all three groups of schools simultaneously: the experimental treatment schools, the experimental control schools, and the non-experimental matched comparison schools. Accordingly, we replace the single treatment status indicator used in previous equations with two group membership indicators: a variable *G1* that equals 1 if the school is in the experimental treatment group and zero 0 otherwise, and a variable *G2* that equals 1 if the school is in the experimental control group and 0 otherwise; membership in the non-experimental comparison group is the omitted category. The longitudinal hybrid model can be expressed as follows:

$$(7) \qquad Y = \gamma_0 + \gamma_1 G1 + \gamma_2 G2 + \sum_{t=2}^{4} \delta_t F_t + \sum_{t=2}^{4} \lambda_t F_t \times G1 + \sum_{t=2}^{4} \theta_t F_t \times G2 + \gamma_3 Z + \beta_1 W + \alpha_1 X + \psi$$

where variables are as defined previously and we omit subscripts for simplicity.

Means and impacts for the first follow-up period can be obtained using the estimated parameters from the first three terms. The estimate of γ_0 represents the regression-adjusted mean outcome in the first follow-up period for matched comparison schools. The estimate of γ_1 represents the difference in the mean outcome between the treatment group and the comparison group in the first follow-up period—that is, the quasi-experimental estimate of the impact of one year of TAP. The estimate of γ_2 represents the difference in the mean outcome between experimental control group and the non-experimental matched comparison group in the first follow-up period. A small, statistically insignificant estimate of γ_2 would suggest that comparison schools and control schools are similar, thus increasing our confidence in the quasi-experimental comparisons. The estimated experimental impact—the difference in the mean outcome between the treatment group and the experimental control group—can be estimated for the first follow-up period as $\gamma_1 - \gamma_2$. Impacts for other follow-up periods and changes in impacts over time can be estimated

through various linear combinations of γ_0, γ_1, γ_2, λ_t, and θ_t. F-tests of hypotheses specified as linear combinations of these parameters will enable us to explore the pattern of impacts.

Subgroup Analysis

Subgroup analysis will be important for revealing whether TAP has greater impacts for some types of schools or teachers than others. However, as discussed in the statistical power section below, the relatively small sample of schools limits our ability to detect meaningful impacts, a difficulty which is exacerbated in subgroup analyses. Furthermore, it is important to exercise caution when interpreting impacts for subgroups defined by variables such as teacher experience levels. Any observation of an impact that differs between novice and experienced teachers does not necessarily imply that novice teachers would have the same impact as the experienced teachers if they had more years of experience. It is possible that teachers with more or less experience differ along any number of unobserved dimensions that could also explain the difference in impacts. In other words, because years of teaching experience are not randomly assigned, we cannot attribute differences in impacts to teacher experience levels. Nevertheless, the subgroup analyses may be instructive as descriptive measures of how the impacts are distributed across schools and teachers.

We will estimate subgroup impacts by adding terms for the interaction between the treatment status indicator and an indicator for membership in the subgroup under consideration. The following is an example of a model that can be used to estimate in a given follow-up period impacts for a subgroup defined based on a teacher characteristic:

$$(8) \qquad Y_{ijk} = \gamma_0 + \gamma_1 T_k + \gamma_2 T \times W_{jk} + \gamma_3 Z_k + \beta_1 W_{jk} + \alpha_1 X_{ijk} + \psi_{ijk}$$

where W is a subgroup membership indicator that equals 1 if teacher *j* belongs to the subgroup and 0 otherwise. For example, if we are interested in the differential impact of TAP on experienced teachers, we can use an indicator that equals 1 if years of experience exceeds a cutoff value and 0 otherwise. The impact of the program on experienced teachers is then estimated by $\gamma_1 + \gamma_2$. An estimate of γ_2 that is statistically significant and positive would suggest that the impact of the program is larger for experienced teachers.

Weights

As discussed in Chapter II, the 2007 lottery used simple random assignment with unequal selection probabilities based on school readiness ratings given by CPS. Impacts will be calculated using sample weights in order to produce estimates for the school of "typical" readiness. Specifically, we will weight schools by the inverse of their selection probability so that schools with a higher probability of being selected into the sample will be given a lower weight.

Weights will also be adjusted to account for interview nonresponse. Although we anticipate high response rates to interviews, surveys, and student tests, data may not be available for some principals, teachers, or students. We can use propensity-scoring

procedures and baseline data to predict the probability that a baseline sample member will respond to a given instrument at follow up. The inverse of these predicted probabilities can then be used as nonresponse weights, giving more weight to respondents whose baseline characteristics are more similar to those of nonrespondents.

Presentation of rResults

The school-level impacts discussed thus far may hide interesting variation in the impacts of TAP across grades. To allow for this possibility, we will present results both disaggregated by grade and aggregated to the school level. Tables IV.1 and IV.2 are shells to illustrate the basic presentation of results. Table IV.1 reports regression-adjusted mean outcomes by treatment status: treatment, control, non-experimental comparison group 1, and non-experimental comparison group 2. Table IV.2 presents the impacts by comparison type, calculated as the difference in the regression-adjusted mean.

Table IV.1. Example of Table Presenting Achievement Gain Scores (Regression-Adjusted)

		TAP	Non-TAP		
Subject	Grade	Treatment	Control	Comparison 1	Comparison 2
Reading/Language Arts	4	xx	xx*	xx*	xx*
	5	xx	xx*	xx*	xx*
	6	xx	xx*	xx*	xx*
	7	xx	xx*	xx*	xx*
	8	xx	xx*	xx*	xx*
	Average	xx	xx*	xx*	xx*
Mathematics	4	xx	xx*	xx*	xx*
	5	xx	xx*	xx*	xx*
	6	xx	xx*	xx*	xx*
	7	xx	xx*	xx*	xx*
	8	xx	xx*	xx*	xx*
	Average	xx	xx*	xx*	xx*
Sample size	Students	nn-nn per grade	nn-nn per grade	nn-nn per grade	nn-nn per grade
	Classrooms	nn-nn per grade	nn-nn per grade	nn-nn per grade	nn-nn per grade

* Difference between non-TAP and TAP score is statistically significant at the 0.05 level (two-sided test).
Note: This table shell can be used for other outcomes, such as teacher turnover, student promotion, etc. This footnote will list the variables used in the regression.

Table IV.2. Example of Table Presenting Impacts of TAP on Achievement

		Impact, by comparison type: TAP (Treatment) versus		
Subject	Grade	Control	Comparison 1	Comparison 2
Reading/Language Arts	4	xx*	xx*	xx*
	5	xx*	xx*	xx*
	6	xx*	xx*	xx*
	7	xx*	xx*	xx*
	8	xx*	xx*	xx*
	Average	xx*	xx*	xx*
Mathematics	4	xx*	xx*	xx*
	5	xx*	xx*	xx*
	6	xx*	xx*	xx*
	7	xx*	xx*	xx*
	8	xx*	xx*	xx*
	Average	xx*	xx*	xx*

* Difference between non-TAP and TAP score is statistically significant at the 0.05 level (two-sided test).
Note: This table shell can be used for other outcomes, such as teacher turnover, student promotion, etc.

Statistical Power

A statistical power analysis demonstrates how well the study's design will be able to distinguish real impacts from chance differences. Our approach to power analysis is to calculate the minimum detectable effect (MDE) for each outcome, given a set of assumptions about sample sizes, other aspects of the design structures, and the variability of the outcomes of interest. The MDE represents the smallest impact, measured in effect size units (that is, as a percentage of the standard deviation of the outcome) that can be detected with high probability. A smaller MDE indicates greater statistical precision.

Table IV.3 shows the MDEs that can be achieved in estimating TAP's impact in a given year as we allow the number of schools analyzed to vary.[8] For example, our calculations suggest that using the eight treatment and eight control schools from a single lottery allows us to detect an impact of 48 percent of a standard deviation. Increasing the sample size by using matched comparison schools or combining schools across lotteries results in greater precision. The statistical power of other analyses we conduct will vary, with the precision of the impact estimates being greater for longitudinal analyses and lower for subgroup analyses. Weighting adjustments made to account for the unequal selection probabilities and interview nonresponse will also affect the standard errors of the estimates and result in slightly larger MDEs than reported here.

[8] In calculating MDEs, we made assumptions typical for estimating the impact of student achievement. A similar analysis can be conducted for other outcomes of interest. The technical assumptions we made are described in the table notes.

Table IV.3. Minimum Detectable Effects (MDE)

	Number of Treatment Schools	Number of Control/Comparison Schools	MDE
Single lottery sample	8	8	0.48
	8	24	0.38
	8	300	0.33
Combined lottery sample	16	16	0.33
	16	48	0.27
	16	300	0.24

Assumptions:
The proportion of total variance in test scores due to between-school variation (intraclass correlation) is 0.15.
50 percent of within-school variance in test scores is explained by covariates (including pretest).
10 percent of between-school variance in test scores is explained by covariates.
15 classrooms per school.
23 students per classroom, with 20 percent attrition/missing data.
Two-tailed hypothesis test conducted with 80 percent power and $\alpha = .10$.

While we have followed common practice and assumed that non-experimental comparison schools can be analyzed equivalently to experimental control schools in calculating statistical power, we note that the MDEs we can detect in analyses that include quasi-experimental comparisons will depend on how well matched the comparison schools are to the treatment schools. If TAP schools and matched non-TAP schools differ on the school characteristics included in the model, the precision of the estimates will decrease. Thus, the MDEs we can achieve may be somewhat higher than those illustrated in Table IV.3.

We note that the MDEs reported in Table IV.3 are higher than desired, suggesting that TAP may produce a positive and meaningful impact in Chicago schools, but that our study design would consider it statistically insignificant (a Type II error, or false negative result). In other words, the finding would be internally valid and descriptive of the particular schools in the study, but not easily generalized to other schools. This is a very common concern with group randomized trials such as the one we are conducting. In our study, the achieved MDEs may be different from the predicted MDEs listed in Table IV.3 for several reasons, some of which were noted above: they will be slightly larger if we account for sample weights and, in the case of the non-experimental estimates, the uncertainty in the estimate of the propensity score; they will be smaller if our estimates of the intraclass correlation (0.15) and R-squared (50 percent at the student level and 10 percent at the school level) turn out to be too conservative.

Another consideration is that group randomized trials (also called cluster randomized trials) produce evidence that often fails to meet the thresholds of statistical significance using classical statistical inference (typically 90 or 95 percent "confidence" level and 80 percent power), but generate meaningful evidence for policymakers using a Bayesian approach to statistical inference. In the Bayesian framework, we begin with a prior belief about the

effectiveness of TAP and update that belief based on evidence generated by the experiment. With Bayesian inference, we can further extract the value of the planned study by incorporating it into existing and future evidence from other trials on TAP's impacts, for example using the method of Blitstein et al. (2005). On balance, we believe that despite the risk of Type II error, using the classical framework is warranted in order to generate new evidence on TAP.

IMPLEMENTATION ANALYSIS

The implementation analysis will synthesize quantitative and qualitative data from TAP, including data from site visits and teacher attitude surveys, as well as MPR teacher surveys and interviews with principals, NIET staff, and CPS staff. Below we describe how we will analyze these data to address the key implementation questions.

The descriptive analysis of the extent to which Chicago public schools faithfully implemented TAP according to the model will draw primarily on program fidelity measures supplied by NIET. Based on our conversations with NIET staff, we expect to obtain fidelity scores on a number of dimensions. We will present tables reporting average scores on each dimension for treatment schools, control schools after they begin implementing TAP, and where possible, TAP schools outside of Chicago. Using data from NIET surveys and other information gathered from interviews with NIET staff, summary tables will also describe ways in which schools tend to deviate from the model when such departures occur. Since treatment and control schools differ in the number of years of TAP implementation, comparing the fidelity of treatment and control schools at a given point in time will suggest how implementation fidelity changes as schools gain experience with the program. Comparisons to other TAP schools across the country will provide context for understanding the results in Chicago. We will also relate the effectiveness of TAP to implementation fidelity when interpreting the main impact findings.

Using richly descriptive qualitative data collected from interviews and program documents, we will seek to identify factors that may facilitate or impede the successful implementation of TAP. Principal interviews will provide important contextual information for each TAP school, such as special policy initiatives operating at the same time as TAP implementation and crises that arose during the school year, as well as data on the principals' perceptions of implementations challenges and strategies attempted to overcome those challenges. Interviews with CPS and NIET staff members will supply an additional set of valuable perspectives on implementation barriers or facilitators and specifics regarding selection, training, and other program processes. These qualitative findings can be synthesized and assessed across schools.

Examining patterns across schools and relating these patterns to differences in school contexts and practices will enable us to draw lessons about promising practices and critical challenges related to implementation. For example, we expect to gather details on program rules and guidelines provided to the intervention schools. In comparing rules developed in Chicago with those of TAP schools in other parts of the country, we can compile simple cross-classification tables to assess whether specific variations in rules are associated with

greater or lesser fidelity to the TAP model. We can also create tabular summaries to highlight factors that are likely to promote or hinder implementation, as identified in our qualitative analysis of Chicago schools.

The implementation analysis will also involve a comparison of how TAP schools differ from practices normally implemented in Chicago public schools. This aspect of the implementation analysis will enable us to characterize the counterfactual condition—that is, the experiences that would have occurred in the absence of TAP. MPR's surveys of teachers in the treatment schools, control schools, and a subset of matched comparison schools will provide data on activities that might be present in both TAP and non-TAP schools, such as mentoring, professional development, and teacher evaluation. For the treatment and control schools, information on these practices will be supplemented with data collected from principal interviews. We will present summary statistics comparing average practices across TAP and non-TAP schools.

In addition to examining descriptive statistics, we can use specific practices as outcomes in regressions analogous to those conducted for the main impact analyses. For example, the teacher survey asks about the frequency and duration of meetings with a mentor. Using a mentoring intensity variable as an outcome, we can calculate regression-adjusted means of mentoring intensity for treatment, control, and comparison schools and assess the extent to which practices in TAP schools differ significantly from activities that occur in non-TAP schools. Such information is critical in interpreting the impact findings.

Reporting and Dissemination

The evaluation will produce five annual reports. This document is the Year 1 report, presenting the refined study design and analysis plan. The Year 2 and Year 3 reports will present findings on implementation and impacts for the first two years of TAP implementation. The Year 4 report will update the findings for the third year of TAP implementation, and will also include findings for the set of schools randomly assigned in the second lottery at the conclusion of their initial year of implementation. The final report will summarize program implementation and impacts across all years for schools from both lotteries. In addition to producing more precise estimates by using schools from both lotteries, this final report will conduct longitudinal analyses to explore the relationships among impacts over time.

In addition to producing the annual reports, we will report periodically during the year to the Joyce Foundation on the status of the evaluation and include interim findings as appropriate. This interim reporting will allow us to identify problems as they arise and develop effective strategies and solutions for overcoming them. It will also provide opportunities to share key results with CPS and the managers of Project REAL so they may benefit from the study and use it to improve the program. Specifically, we plan to conduct two in-person briefings per year for stakeholders in Chicago, one in July and one in December. The July briefings will focus on implementation. In 2008 and 2010 the July briefing will include results from the teacher interviews and staff interviews. The July 2009 briefing will discuss the results of school recruitment and random assignment. The

December briefings will include findings from the analysis of student test score and teacher mobility data provided by CPS, as well as principal interviews. The dates and audiences for the briefings will be revised as necessary in consultation with key stakeholders: the Joyce Foundation, Project REAL staff, and NIET.

Dissemination of the study's findings is critical for maximizing the impact on policy. We will prepare versions of the final report that will serve two audiences. One version will be technical, resulting in manuscripts suitable for publication in journals such as the *Journal of Human Resources*, *Economics of Education Review*, *Journal of Policy Analysis and Management*, and *Education Finance and Policy*. The other version will be nontechnical, suitable for general or policy audiences and will be posted on the Mathematica website. We will use a special issue brief format to help disseminate the findings to a wide audience. MPR issue briefs use advanced graphics and printing to present concise summaries of longer reports.

To further strengthen these dissemination efforts, the evaluation team will provide occasional briefings to policymakers, stakeholders, and researchers. These audiences include CPS staff, NIET staff, funders of the project, and attendees of research conferences and seminars such as the American Educational Research Association, the American Education Finance Association, and workshops held by the University of Chicago's Department of Economics and the Harris School of Public Policy. Table IV.4 presents a detailed evaluation schedule.

Table IV.4. Project Reporting and Dissemination Timeline

Product	Date Delivered
Draft Annual Report – Refined Design and Analysis Plan	July 31, 2007
Revised First Annual Report	November 15, 2007
In-person briefing, Year 1 implementation[b]	July 2008
Second Annual Report –Year 1 TAP Impacts	November 14, 2008
MPR Issue Brief	November 28, 2008[a]
In-person briefing, Year 1 impacts	December 2008
In-person briefing, Year 2 implementation	July 2009
In-person briefing, Year 2 impacts	December 2009
Third Annual Report –Year 2 TAP Impacts	January 29, 2010
In-person briefing, Year 3 implementation	July 2010
Fourth Annual Report –Year 3 TAP Impacts and Year 1 TAP Impacts for Second Lottery Schools	November 1, 2010
In-person brefing, Year 3 impacts	December 2010
In-person briefing, Year 4 impacts preview	October 2011
MPR Issue Brief	November 1, 2011
Fifth Annual Report – Impacts and Longitudinal Analyses for All Available Years	November 1, 2011

[a] In-person briefings will be offered to core stakeholders (Project REAL staff, NIET staff, and the Joyce Foundation) at each occasion. As deemed appropriate by CPS, briefings will be offered to other stakeholders (district and school building staff) at the same time or in addition.

REFERENCES

Azordegan, Jennifer, Patrick Byrnett, Kelsey Campbell, Josh Greenman, and Tricia Coulter. "Diversifying Teacher Compensation." Issue Paper. Washington, DC: Education Commission of the States and the Teaching Commission, December 2005.

Blitstein, J.L, D.M. Murray, P.J. Hannan, and W.R. Shadish. "Increasing the Degrees of Freedom in Future Group Randomized Trials: The df* Approach." *Evaluation Review*, vol. 29, 2005, pp. 268-286.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina." Cambridge, MA: National Bureau of Economic Research, June 2006.

Cook, Thomas, and Donald Campbell. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Hopewell, NJ: Houghton Mifflin, 1979.

Glazerman, Steven. "Teacher Compensation Reform: Promising Strategies and Feasible Methods to Rigorously Study Them." Washington, DC: Mathematica Policy Research, December 2004.

Glazerman, Steven, Tim Silva, Nii Addy, Sarah Avellar, Jeffrey Max, Allison McKie, Brenda Natzke, Michael Puma, Patrick Wolf, and Rachel Ungerer Greszler. "Options for Studying Teacher Pay Reform Using Natural Experiments." Washington, DC: Mathematica Policy Research, March 2006.

Gordon, Robert, Thomas Kane, and Douglas Staiger. "Identifying Effective Teachers Using Performance on the Job." Washington, DC: The Hamilton Project, Brookings Institution, June 2006.

Hanushek, Eric, and Steven Rivkin. "How to Improve the Supply of High Quality Teachers." Brookings Papers on Education Policy. Washington, DC: Brookings Institution, May 2003.

Hassel, B.C. "Better Pay for Better Teaching: Making Teacher Compensation Pay Off in the Age of Accountability." Washington, DC: Progressive Policy Institute, May 2002.

Hatry, Harry P., John M. Greiner, and Brenda G. Ashford. *Issues and Case Studies in Teacher Incentive Plans*. Second Edition. Washington, DC: The Urban Institute Press, 1994.

McCaffrey, Daniel, J.R. Lockwood, Daniel Koretz, and Laura Hamilton. "Evaluating Value-Added Models for Teacher Accountability." Santa Monica, CA: RAND Corporation, 2003.

Murnane, Richard, and David K. Cohen. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review*, vol. 56, no. 1, February 1986, pp. 1-17.

Myers, David, and Mark Dynarski. "Random Assignment in Program Evaluation and Intervention Research: Questions and Answers." Washington, DC: U.S. Department of Education, Institute of Educational Sciences, National Center for Education Evaluation and Regional Assistance, June 2003.

Podgursky, Michael. "The Single Salary Schedule for Teachers in K-12 Public Schools." Discussion paper prepared for the Center for Reform of School Systems. University of Missouri-Columbia, August 2002.

Sanders, William, and June Rivers. "Cumulative and Residual Effects of Teachers on Future Student Academic Achievement." Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, November 1996.

Sanders, William, and Sandra Horn. "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research." *Journal of Personnel Evaluation in Education*, vol. 12, no. 3, September 1998, pp. 247-256.

Schacter, John, Tamara Schiff, Yeow Meng Thum, Cheryl Fagnano, Micheline Bendotti, Lew Solmon, Kimberly Firetag, and Lowell Milken. "The Impact of the Teacher Advancement Program on Student Achievement, Teacher Attitudes, and Job Satisfaction." Santa Monica, CA: Milken Family Foundation, November 2002.

Schacter, John, Yeow Meng Thum, Daren Reifsneider, and Tamara Schiff. "The Teacher Advancement Program Report Two: Year Three Results from Arizona and Year One Results from South Carolina TAP Schools." Santa Monica, CA: Milken Family Foundation, March 2004.

Solomon, Lewis, J. Todd White, Donna Cohen, and Deborah Woo. "The Effectiveness of the Teacher Advancement Program." Santa Monica, CA: National Institute on Effectiveness in Teaching, April 2007.

TeacherSolutions. "Performance-Pay for Teachers: Designing a System That Students Deserve." Hillsborough, NC: National Center for Teaching Quality, March 2007.

Working Group on Teacher Quality. "Creating a Successful Performance Compensation System for Educators." Santa Monica, CA: National Institute for Excellence in Teaching, July 2007.

APPENDIX A

DRAFT TEACHER QUESTIONNAIRE

DRAFT TEACHER SURVEY FOR EVALUATION OF CHICAGO TAP

[Introductory text with endorsements and burden estimate will go here]

SECTION A. EDUCATIONAL AND PROFESSIONAL BACKGROUND

EDUCATION

A1. Please describe your postsecondary degrees in the chart below.

A. YEAR AWARDED	B. TYPE OF DEGREE	C. NAME , CITY, AND STATE OF INSTITUTION	D. MAJOR FIELD OF STUDY	E. MINOR FIELD OF STUDY
$$$$$$$$	1 ☐ Associate's 2 ☐ Bachelor's			
$$$$$$$$	1 ☐ Bachelor's 2 ☐ Master's 3 ☐ Other *(SPECIFY):* _____			
$$$$$$$$	1 ☐ Bachelor's 2 ☐ Master's 3 ☐ Other *(SPECIFY):* _____			

A2. Are you currently working toward an advanced degree (for example, Master's, Ed.D., or Ph.D.), additional credits or certificates?

1 ☐ Yes → 1 ☐ Degree: _____

0 ☐ No → 2 ☐ Additional Credits: _____

3 ☐ Certificates: _____

GO TO A3

a. Name, city, and state of college or university

_____

b. Major Field of Study

_____

CERTIFICATION

A3. Do you have a teaching certificate issued by the Illinois State Board of Education?

1 ☐ Yes → ***GO TO A4***

0 ☐ No → **a. Are you currently pursuing state certification?**

1 ☐ Yes
0 ☐ No
→ ***GO TO A9***

A4. The list below shows the types of certification that teachers in Chicago may obtain. Indicate the certificate, or certificates, that you currently hold by placing a check in the appropriate box/es.

CODE	DESCRIPTION	
2	Early Childhood Teaching	☐
3	Elementary Certificate	☐
4	Early Childhood	☐
5	Provisional Early Childhood	☐
9	Secondary Certificate	☐
10	Special Certificate	☐
22	Alternative Elementary Teaching	☐
23	Provisional Alt Elementary	☐
24	Alternative Secondary Teaching	☐
25	Provisional Alt Secondary	☐
26	Provisional Alt Administrative	☐
27	Provisional Alt Special	☐
29	Transitional Bilingual Certificate	☐
30	Provisional Elementary	☐
31	Provisional Secondary	☐
33	Provisional Special	☐
34	Provisional Vocational Teaching	☐
38	Resident Teacher Certification	☐
39	Substitute-90 Days Teaching	☐
43	Provisional Alt Early Childhood	☐
50	Special Visiting International Teacher	☐
73	School Service Personnel (Includes endorsements: School Nurse, School Psychologist, School Social Worker, School Counselor and Speech Language Pathologist-Non Teaching)	☐
74	Provisional School Service Personnel	☐
75	Administration	☐
	Other (specify) _____	☐

A5. When did you receive your ORIGINAL teaching certificate?

$$$$$$$$ / $$$$$$$$$$$$$$$$
MONTH YEAR

A6. When did you obtain a renewal of your teaching certificate?

NOTE: IF YOU RECEIVED YOUR ORIGINAL CERTIFICATE WITHIN THE LAST 5 YEARS, CHECK THE BOX AND SKIP TO QUESTION A7 .....☐

$$$$$$$$ / $$$$$$$$$$$$$$$$
MONTH YEAR

A7. Which of the following statements best describes how you earned your teaching certificate? *CHECK ONE BOX ONLY.*

1 ☐ In **a traditional teacher certification program** as part of a bachelor's degree

2 ☐ In **a traditional teacher certification program** as a "5th year" or master's degree

3 ☐ As part of an **alternative teacher certification program**

4 ☐ Other *(SPECIFY):*

_____

A8. Do you have, or are you currently pursuing, advanced professional certification, such as National Board Certification or related credential? *CHECK ONE BOX ONLY.*

1 ☐ I have National Board certification

→ If yes, in what discipline is this certification? _____

2 ☐ I am pursuing National Board certification

→ If yes, in what discipline? _____

3 ☐ I have another advanced teacher certification

(SPECIFY THE CREDENTIAL): _____

4 ☐ I am pursuing another advanced teacher certification

(SPECIFY THE CREDENTIAL): _____

5 ☐ None of the above

YOUR CURRENT TEACHING POSITION

A9. In what grade(s) are the students you currently teach THIS school year? *CIRCLE ALL THAT APPLY.*

K 1 2 3 4 5 6 7 8 9 10 11 12

A10. Which of the following statements best describes your teaching assignment THIS school year? *CHECK ONE BOX FOR THE RESPONSE THAT BEST APPLIES.*

1 ☐ **Self-Contained Instruction —** I instruct the same group of students most or all of the day in multiple subjects.

2 ☐ **Departmentalized Instruction —** I instruct several classes of different students most or all of the day in one or more academic subjects (such as reading and language arts, math, algebra, science, or social studies).

(SPECIFY SUBJECTS): _____

3 ☐ **Enrichment Instruction —** I instruct more than one class of students most or all of the day in one enrichment, or elective, subject (such as art, music, physical education or computer skills).

SPECIFY SUBJECTS): _____

4 ☐ **"Pull-out" Class —** I instruct selected students released from their regular classes in specific skills or to address specific needs (such as gifted and talented, special education, reading, English as a Second Language)

(SPECIFY SPECIFIC SKILLS, NEEDS, OR CURRICULUM AREA): _____

5 ☐ **Team Teaching —** I am one of two or more teachers, in the same class, at the same time, and we are jointly responsible for teaching a single group of students.

6 ☐ Other type of teaching assignment *(SPECIFY):* _____

A11. How much time do you work as a teacher at THIS school? *CHECK ONE BOX ONLY.*

1 ☐ Full time

2 ☐ Less than full time, but more than half time

3 ☐ Half time

4 ☐ Less than half time, but more than quarter time

PREVIOUS TEACHING EXPERIENCES

A12. Complete the table below indicating your teaching positions over the last 3 years. *INCLUDE TIME SPENT TEACHING BOTH FULL- AND PART-TIME. INDICATE YOUR TEACHING ASSIGNMENT OR POSITION IN COLUMN A, AND THE AGENCY FOR WHICH YOU WORKED IN COLUMN B, USING THE CATEGORIES LISTED BELOW. DO NOT INCLUDE TIME SPENT AS A STUDENT TEACHER.*

Assignment

1 = Head classroom teacher
2 = Assistant teacher
3 = Teacher aide
4 = Long-term substitute teacher
5 = Other *(SPECIFY IN CHART BELOW)*

Agency

1 = Chicago Public Schools
2 = Regular public school other than CPS
3 = Charter School outside of Chicago
4 = Charter School in Chicago
5 = Other

	A.	B.	C.	D.	E.
SCHOOL YEAR	ASSIGNMENT	AGENCY	SCHOOL NAME	GRADE(S)	REASON FOR LEAVING *(WRITE "NA" IF YOU HAVE NOT LEFT)*
a. 2006-07					
b. 2005-06					
c. 2004-05					

A13. Prior to 2004-05, how many years did you work as a head classroom teacher (do not include time spent as an assistant teacher, teacher aide, student teacher, or substitute teacher)?

$$$$$$ YEARS *IF "0" SKIP TO A15.*

A14. Complete the table below indicating your teaching experience as a HEAD CLASSROOM teacher before 2004-05. *FOR EACH SCHOOL TYPE, INDICATE WHETHER YOU HAVE TAUGHT IN THE SETTING, AND IF SO THE SCHOOL YEAR(S) THAT YOU WERE EMPLOYED, THE GRADE(S) TAUGHT, AND SUBJECTS (IF YOU WERE A SUBJECT SPECIALIST).*

	A.	B.	C.	D.
SCHOOL TYPE	TAUGHT? YES NO	YEAR(S)	GRADE(S)	SUBJECT(S)
a. Chicago Public School	1☐ 0☐			
b. Charter School/ Other District	1☐ 0☐			
c. Private School	1☐ 0☐			

A15. Before 2004-05, how many years did you work as an assistant teacher, teacher aide, or substitute teacher?

$$$$$$$$ YEARS *(IF "0" SKIP TO SECTION B ON PAGE 7.)*

A16. Complete the table below indicating your teaching experience as an assistant teacher, teacher aide, or substitute teacher before 2004-05. *FOR EACH SCHOOL TYPE, INDICATE WHETHER YOU HAVE TAUGHT IN THE SETTING AND, IF SO, YOUR BEST RECOLLECTION OF THE SCHOOL YEAR(S) THAT YOU WERE EMPLOYED, AND THE GRADE(S) TAUGHT.*

	A.	B.	C.
SCHOOL TYPE	TAUGHT? YES NO	YEAR(S)	GRADE(S)
a. Chicago Public School	1☐ 0☐		
b. Charter School /Other District	1☐ 0☐		
c. Private School	1☐ 0☐		

A17. Before your first teaching position, did you student teach?

1☐ Yes → **a. How many weeks?** ..... $$$$$$$$ WEEKS

0☐ No

SECTION B. SUPPORT YOU RECEIVE TO IMPROVE YOUR TEACHING

B1. How much time have you participated in formal professional development activities addressing the topics listed below since July 2007? *INCLUDE COURSES YOU HAVE TAKEN FOR RECERTIFICATION OR ADVANCED CERTIFICATION, WORKSHOPS SPONSORED BY YOUR DISTRICT, CONFERENCES, OR OTHER TRAINING THAT IS RELEVANT TO YOUR TEACHING. CHECK ONE BOX IN EACH ROW.*

PROFESSIONAL DEVELOPMENT ACTIVITY	TIME SINCE JULY 2007				
	NO TIME	LESS THAN 1/2 HOUR	1/2 TO 1 HOUR	1 TO 2 HOURS	MORE THAN 2 HOURS
a. Lesson planning	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
b. Learning to use best-practice instructional techniques	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
c. Reviewing student performance data to inform instruction	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
d. Improving classroom management skills	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
e. Increasing your subject matter or content area knowledge	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
f. Aligning curriculum to state standards and/or vertically across grade levels	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
g. Aligning local or teacher-developed curriculum assessment to state standards	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
h. Preparing students for standardized tests	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
i. Student motivation and engagement	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
j. Differentiated instruction (i.e., varied approaches to instruction in response to student differences in readiness and learning needs)	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
k. Using computers to support instruction	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
l. Strategies for teaching literacy (reading; Language Arts)	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
m. Strategies for teaching math	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
PERHAPS MAKE ITEMS IN THIS QUESTION MATCH THOSE IN Q C9.					
l. Other *(SPECIFY)* _____	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐

B2. On average, would you characterize the usefulness of the professional development activities you attended since July 2007 as . . .

MARK ONLY ONE BOX

1 ☐ Not at all useful to your teaching?
2 ☐ Mostly not useful to your teaching?
3 ☐ Mostly useful to your teaching?
4 ☐ Very useful to your teaching?

B3. Compared to other professional development activities in which you have participated, would you characterize those you have attended since July 2007 as . . .

MARK ONLY ONE BOX

1 ☐ Much less satisfactory?
2 ☐ Mostly not as satisfactory?
3 ☐ Equally satisfactory?
4 ☐ Mostly more satisfactory?
5 ☐ Much more satisfactory?

B4. To what extent have you been able to implement in your teaching what you have learned in the professional development activities in which you have participated since July 2007.

MARK ONLY ONE BOX

1 ☐ Not at all
2 ☐ Slightly
3 ☐ Somewhat
4 ☐ A great deal

B5. Associated with any of the above professional development activities, did you receive the following?

MARK YES OR NO FOR EACH	YES	NO
a. Release time from teaching	1 ☐	0 ☐
b. Scheduled time in the contract year for professional development	1 ☐	0 ☐
c. Stipend for professional development activities that took place outside regular work hours	1 ☐	0 ☐
d. Full or partial reimbursement of college tuition	1 ☐	0 ☐
e. Reimbursement for conference or workshop fees	1 ☐	0 ☐
f. Reimbursement for travel and/or daily expenses	1 ☐	0 ☐
g. Credits toward re-certification or advanced certification	1 ☐	0 ☐
h. Increase in salary or other pay increases	1 ☐	0 ☐
i. Recognition or higher ratings on an annual teacher evaluation	1 ☐	0 ☐

B6. During the current school year, how many times were you . . .

IN EACH ROW, CHECK ONE BOX ONLY	NEVER	1 TIME	2 TIMES	3 TIMES	4 TIMES	5 OR MORE TIMES
a. Observed teaching your class by your principal or assistant/vice principal?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
b. Observed teaching your class by a mentor, coach, or master teacher?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
c. Given feedback on your teaching as part of a formal evaluation process?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
d. Given feedback on your teaching (not as part of a formal evaluation process)?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
e. Given feedback on your lesson plans?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐

EXPERIENCES WITH A MENTOR, COACH, OR MASTER TEACHER

B7. Is there currently a person or persons, such as a mentor, coach, master teacher, or other school or district leader, who provides professional advise and direct assistance to you in your teaching duties? *CHECK ALL THAT APPLY.*

1 ☐ Yes →
- 1 ☐ Mentor
- 2 ☐ Coach
- 3 ☐ Master Teacher
- 4 ☐ Other *(SPECIFY)*_____

0 ☐ No →

GO TO SECTION C ON PAGE 12

IF YOU ARE WORKING WITH MORE THAN ONE MENTOR, COACH, OR MASTER TEACHER, PLEASE RESPOND TO THE FOLLOWING QUESTIONS FOR THE ONE WITH WHOM YOU SPEND THE MOST TIME. INDICATE THE TITLE OF THE PERSON IN THE SPACE PROVIDED.

B8. Which of the following statements BEST describes your _____?

CHECK ALL THAT APPLY.

Is a full-time teacher	1 ☐
Is a part-time teacher	2 ☐
Works in your school only	3 ☐
Works in more than one school	4 ☐
Has some release time from teaching his/her own classroom	5 ☐
Has no responsibility for teaching a classroom	6 ☐
Is a school-based administrator *(SPECIFY TITLE):* _____	7 ☐
Is a school-based specialist *(SPECIFY TITLE):* _____	8 ☐
Is a district administrator *(SPECIFY TITLE):* _____	9 ☐
Is a district specialist *(SPECIFY TITLE) :*_____	10 ☐
Is someone from a teacher licensing, certification, or preparation program (including universities and colleges)	11 ☐
Other *(SPECIFY):* _____	12 ☐

B9. On average, how often do you have scheduled meetings with this person? *CHECK ONE BOX ONLY.*

Daily (5 or more times per week)	1 ☐
2-4 times per week	2 ☐
Once a week	3 ☐
2-3 times per month	4 ☐
Once a month	5 ☐
Every other month	6 ☐
Fewer than 5 times per year	7 ☐
Other *(SPECIFY):* _____	8 ☐

B10. On average, how long were each of these meetings with this person? *WRITE YOUR ESTIMATE IN TERMS OF TOTAL HOURS AND MINUTES IN THE BOXES BELOW.*

$$$$$$$$ HOURS, $$$$$$$$ MINUTES PER MEETING

B11. In addition to the scheduled meetings referenced in Question B8 above, how much informal (not scheduled) contact did you have with this person each MONTH? *WRITE YOUR ESTIMATE IN TERMS OF TOTAL HOURS AND MINUTES IN THE BOXES BELOW.*

$$$$$$$$ HOURS, $$$$$$$$ MINUTES PER MONTH

B12. During the most recent full week of teaching, how much scheduled time did this person spend . . .

IN EACH ROW, CHECK ONE BOX ONLY	NO TIME	LESS THAN 1/2 HOUR	1/2 TO 1 HOUR	1 TO 2 HOURS	MORE THAN 2 HOURS
a. Observing your teaching?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
b. Meeting with you on a one-to-one basis?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
c. Meeting with you together with other teachers?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
d. Modeling a lesson?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
e. Co-teaching a lesson?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐

B13. How would you rate the quality of the feedback that you received from these interactions with this person with respect to its helpfulness to your teaching?

MARK ONLY ONE BOX

1 ☐ Not at all helpful
2 ☐ Slightly helpful
3 ☐ Somewhat helpful
4 ☐ Very helpful

SECTION C. SUPPORT AND LEADERSHIP YOU PROVIDE

Questions C1 through C10 ask about your experiences as a mentor teacher. Questions C11 through C13 ask about other leadership roles and responsibilities you may have.

C1. Do you provide formal mentoring services to other teachers at your school?

1 ☐ Yes

0 ☐ No → ***GO TO C11***

C2. Are you given release time from your regular instructional duties to perform the tasks related to your position as a mentor?

1 ☐ Yes → How many hours in a typical week? .$$$$$$ HOURS

0 ☐ No

C3. On average, how many hours OUTSIDE of your specified contracted hours do you spend each week engaged in activities related to your position as a mentor?

.$$$$$$ HOURS EACH WEEK *(IF NONE, ENTER 0.)*

C4. How many teachers are you currently mentoring on an ongoing basis?

.$$$$$$ TEACHERS

C5. On average, how often do you have scheduled meetings with EACH teacher you are mentoring? *CHECK ONE BOX ONLY.*

- Daily (5 or more times per week) ☐ 1
- 2-4 times per week ☐ 2
- Once a week ☐ 3
- 2-3 times per month ☐ 4
- Once a month ☐ 5
- Every other month ☐ 6
- Fewer than 5 times per year ☐ 7
- Varies by teacher according to need ☐ 8
- Other *(SPECIFY):* _____ ☐ 9

C6. On average, how long were these meetings with each teacher? *WRITE YOUR ESTIMATE IN TERMS OF TOTAL HOURS AND MINUTES IN THE BOXES BELOW.*

$$$$$$$$ HOURS, $$$$$$$$ MINUTES PER MEETING

C7. In addition to the scheduled meetings referenced in Question C5, how much informal (not scheduled) contact did you have collectively with the all of the teachers you were mentoring each WEEK? *WRITE YOUR ESTIMATE IN TERMS OF TOTAL HOURS AND MINUTES IN THE BOXES BELOW.*

$$$$$$$$ HOURS, $$$$$$$$ MINUTES PER WEEK

C8. Which of the following activities are part of your responsibilities as a mentor teacher? *CHECK ALL THAT APPLY.*

- Observing teachers teach ☐ 1
- Meeting with teachers on a one-to-one basis ☐ 2
- Meeting with teachers together in small groups or clusters ☐ 3
- Modeling lessons ☐ 4
- Co-teaching lessons ☐ 5
- Preparing written evaluations of teachers ☐ 6
- Other *(SPECIFY):* _____ ☐ 7

C9. Which of the following content areas are covered in your mentoring activities?
CHECK ALL THAT APPLY.

Teacher reading/language arts	1 ☐
Teaching math	2 ☐
Teaching subjects other than reading or math *(SPECIFY):* _____	3 ☐
Teaching to meet state or district standards	4 ☐
Setting instructional goals and ways to achieve them	5 ☐
Assessing students	6 ☐
Using student assessments to inform teaching	7 ☐
Preparing lesson plans or other instructional activities	8 ☐
Teaching children with varying levels of achievement/ability	9 ☐
Using multiple instructional strategies/techniques to teach students	10 ☐
Selecting or adapting curriculum materials	11 ☐
Reflecting on teachers' instructional practices	12 ☐
Motivating students	13 ☐
Other *(SPECIFY):* _____	14 ☐

C10. During the most recent full week of teaching, how much scheduled time did you spend . . .

IN EACH ROW, CHECK ONE BOX ONLY	NO TIME	LESS THAN 1/2 HOUR	1/2 TO 1 HOUR	1 TO 2 HOURS	MORE THAN 2 HOURS
a. Observing teachers teach?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
b. Meeting with teachers on a one-to-one basis?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
c. Meeting with teachers together in small groups or clusters?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
d. Modeling a lesson?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
e. Co-teaching a lesson?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
f. Preparing written evaluations of teachers	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐
g. Other *(SPECIFY)*: _____	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐

C11. Do you have other leadership roles or responsibilities, or serve on any policy-making committees, in the school?

1 ☐ Yes

0 ☐ No → ***GO TO SECTION D***

C12. Indicate whether each role or responsibility below applies to you (column A). If "yes", indicate whether you are given release time to perform the tasks required of the position (column B), and/or are given monetary compensation for the position.

In each row, check one box in Column A. If you answer Yes in Column A, complete Columns B and C for that row.	A. *Applies to you?*		B. *Release time?*		C. *Compensation?*	
Roles and Responsibilities	Yes	No	Yes	No	Yes	No
Leadership Roles						
a. Master teacher	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
b. Supervising teacher	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
c. Department head or chair	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
d. Grade-level lead teacher	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
e. Lead curriculum specialist	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
f. Site-based management team	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
g. School improvement team	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
h. School-wide committee or task force	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
i. District-wide committee or task force	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
j. Other *(Specify)*: _____	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
k. Other *(Specify)*: _____	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
l. Other *(Specify)*: _____	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
Responsibilities						
m. Setting school policies	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
n. Selecting or developing curriculum	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
o. Reviewing and selecting text books	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
p. Input on improving facilities and technology	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
q. Providing in-services or professional development activities to other teachers	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
r. Developing standards for curriculum and assessments	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
s. Evaluating teachers	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
t. Hiring new teachers	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
u. Other *(Specify)*: _____	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
v. Other *(Specify)*: _____	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐
w. Other *(Specify)*: _____	1 ☐	0 ☐	1 ☐	0 ☐	1 ☐	0 ☐

C13. Associated with any of these professional development activities, did you receive the following?

MARK YES OR NO FOR EACH	YES	NO
a. Credit toward recertification or advanced certification.....	1 ☐	0 ☐
b. Increase in salary or other pay increases.....	1 ☐	0 ☐
c. Recognition or higher ratings on an annual teacher evaluation.....	1 ☐	0 ☐
d. Other *(SPECIFY):* _____	1 ☐	0 ☐

SECTION D. COMPENSATION

The following questions refer to your before-tax earnings from teaching and other employment. Consider the current school year to run from July 1, 2007 to June 30, 2008.

D1. During the current school year, what is your academic-year, base teaching salary?

$ $$$$$$$$$, $$$$$$$$$. $$$$$

D2. Does your base teaching salary include additional compensation for the leadership roles and responsibilities you have taken on this year?

1 ☐ Yes 0 ☐ No

D3. During the current school year, do you expect to receive any additional compensation from the school system for the leadership roles and responsibilities you have taken on this year?

EXCLUDE ANY COMPENSATION FOR COACHING SPORTS TEAMS, LEADING FIELD TRIPS, OR ADVISING STUDENT GROUPS.

1 ☐ Yes → **a. How much? $** $$$$$$, $$$$$$$$$. $$$$$

0 ☐ No

D4. During the current school year, have you earned, or do you expect to earn, additional compensation from working in any job OUTSIDE this school system?

1 ☐ Yes → **a. How much? $** $$$$$$, $$$$$$$$$. $$$$$

0 ☐ No

D5. During the current school year, will you be eligible to receive additional compensation based on the following?

MARK YES OR NO FOR EACH.	YES	NO
a. How well you perform in your position at the school	1 ☐	0 ☐
b. The quality of your instructional performance	1 ☐	0 ☐
c. The achievement growth of the students you teach	1 ☐	0 ☐
d. The subject matter that you teach	1 ☐	0 ☐
e. A particular student population that you work with	1 ☐	0 ☐
f. Other *(SPECIFY):* _____	1 ☐	0 ☐

D6. If you answered "Yes" to any of the items in Question D5, what is your best estimate of the amount of compensation that you will receive?

$ $$$$$$, $$$$$$$$$. $$$$$

D7. How confident are you that you will receive at least this much additional compensation? *WRITE IN ANY AMOUNT BETWEEN 0 AND 100 PERCENT.*

_____ % CONFIDENT

SECTION E. SATISFACTION

E1. How much do you disagree or agree with the following statements about the principal at this school?

IN EACH ROW, CHECK ONE BOX ONLY	STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
a. The principal at this school is strongly committed to shared decision-making	1 ☐	2 ☐	3 ☐	4 ☐
b. The principal at this school works to create a sense of community in the school	1 ☐	2 ☐	3 ☐	4 ☐
c. The principal at this school promotes parent and community involvement in the school	1 ☐	2 ☐	3 ☐	4 ☐
d. The principal at this school supports and encourages teachers to take risks	1 ☐	2 ☐	3 ☐	4 ☐
e. The principal at this school is willing to make changes	1 ☐	2 ☐	3 ☐	4 ☐
f. Most changes introduced at this school receive strong support from the principal	1 ☐	2 ☐	3 ☐	4 ☐
g. The principal at this school encourages teachers to try new methods of instruction	1 ☐	2 ☐	3 ☐	4 ☐

E2. At this time, how satisfied are you with EACH of the following aspects of teaching at this school?

IN EACH ROW, CHECK ONE BOX ONLY	VERY DISSATISFIED	SOMEWHAT DISSATISFIED	SOMEWHAT SATISFIED	VERY SATISFIED
a. Support from administration for teachers	1 ☐	2 ☐	3 ☐	4 ☐
b. Your input into school policies and practices	1 ☐	2 ☐	3 ☐	4 ☐
c. Autonomy or control over your own classroom	1 ☐	2 ☐	3 ☐	4 ☐
d. Opportunities for professional development	1 ☐	2 ☐	3 ☐	4 ☐
e. Professional caliber of colleagues	1 ☐	2 ☐	3 ☐	4 ☐
f. Supportive atmosphere among faculty/collaboration with colleagues	1 ☐	2 ☐	3 ☐	4 ☐
g. Salary and benefits	1 ☐	2 ☐	3 ☐	4 ☐
h. Opportunities for teacher leadership roles and responsibilities	1 ☐	2 ☐	3 ☐	4 ☐
g. School policies	1 ☐	2 ☐	3 ☐	4 ☐

SECTION F. BACKGROUND

F1. Are you male or female?

1 ☐ Male

2 ☐ Female

F2. Are you of Hispanic or Latino origin?

1 ☐ Yes

0 ☐ No

F3. How do you describe yourself? (*SELECT ONE OR MORE*)

1 ☐ American Indian or Alaska Native

2 ☐ Asian

3 ☐ Black or African American

4 ☐ Native Hawaiian or Other Pacific Islander

5 ☐ White

F4. What is your year of birth?

$$$$$$$$$$$$$$$$ YEAR

F5. Did you ever attend the Chicago Public Schools as a student?

1 ☐ Yes

0 ☐ No

CONTACT INFORMATION

Please provide your contact information and the best time to reach you in case we have questions about your responses.

MR./MS./DR. FIRST NAME LAST NAME

STREET APT. NUMBER

CITY STATE ZIP

E-MAIL ADDRESS

( ) PHONE NUMBER (INCLUDE AREA CODE)

BEST TIME TO REACH YOU

THANK YOU FOR COMPLETING THIS SURVEY.