Abstract

Title: Learning Leadership: Kernel Routines for Instructional Improvement
Center: National Center for Education Research
Year of Award: 2010
Principal Investigator: James Spillane
Grantee: Northwestern University
Program: Education Leadership
Goal: Efficacy
Award Period: 4 years
Co-PI: Brian Junker, Carnegie Mellon University; Richard Correnti, University of Pittsburgh

DESCRIPTION

Purpose:
The Learning Walk® routine, developed by the Institute for Learning (IFL) at the University of Pittsburgh, is a form of “walkthrough” or “learning walk” practice in which school leaders conduct brief classroom visits on a regular basis for the purpose of observing classroom instruction. The purpose of the proposed research is to assess the efficacy of the Learning Walk routine, structured by IFL as a strategy for developing school leaders that centers on developing leadership practice through the structured and scaffolded implementation of a kernel organizational routine.

Project Activities:
The Learning Walk is a kernel routine; that is, an organizational routine that has the potential for long-term transformation of school practice by seeding and propagating new forms of professional activity firmly tied to the school’s core work. The hallmarks of a kernel routine approach to school leadership development include: a) developing leadership practice by redesigning the school organization; b) developing leadership practice in situ – in actual schools and on the job; d) developing leadership practice via leadership teams in a school, not just the principal; and e) developing leadership practice firmly anchored in classroom teaching and student learning activities. The research team will conduct an experimental study in which one half of a sample of urban elementary schools are randomly assigned to receive The Learning Walk routine (the treatment group) and the other half (the control group) are assigned to conduct “business-as-usual” according to school and district practice.

Products: Products from this project include published reports on the efficacy of an organizational routine (walkthrough intervention) on school practice and student achievement.

Setting: The setting for this efficacy trial consists of elementary schools within an urban school district.
**Population:** The population will be drawn from elementary schools in an urban school district in which the majority of the students are poor and minority.

**Intervention:** *The Learning Walk routine* is a highly structured set of activities for the observation and interpretation of teaching and learning based on the IFL's Principles of Learning. The *LW* routine and its accompanying training by IFL involve intensive practice-based learning, focused on literacy, for school leaders comprised of a team of the principal and three instructional leaders per school. School leadership teams in the elementary schools randomly assigned to the treatment group will be trained in *The Learning Walk* kernel routine by IFL and supported during its implementation. These treatment school leadership teams will undergo five full-days and five half-days of intensive training. Principals will receive an additional three full-day and five half-day training sessions. Learning walks will occur with a frequency of at least one per week. The leadership teams are expected to conduct a minimum of 12 learning walks in treatment schools, rotating classrooms until every teacher has been visited and has participated in a learning walk at least once.

**Research Design and Methods:**
The research team will conduct an experimental study in which one half of a sample of urban elementary schools are randomly assigned to receive *The Learning Walk routine* (the treatment group) and the other half (the control group) are assigned to conduct “business-as-usual” according to school and district practice. Researchers will examine whether or not the intervention produces changes in school practices (e.g. collaborative routines, classroom instruction, etc.) and student achievement. While the intervention centers on literacy, researchers will analyze mathematics achievement in addition to reading and writing in order to examine whether the intervention transfers to mathematics achievement. Researchers will also conduct a set of embedded case studies to examine how school-level conditions mediate the observed effects on school practice, classroom teaching, and student learning in order to better understand how the intervention enables change.

**Control Condition:** Schools randomly assigned to the control group will conduct "business as usual" using existing school and district practices.

**Key Measures:** The key outcomes of this study are school practice and student achievement. School practice includes measures of collaborative routines (around teaching), expertise of leaders in judging and critiquing classroom instruction, frequency of staff interactions, level of influence of staff on other staff, the number of ties staff have with other staff, school norms (collective responsibility, collaboration, trust, innovation), classroom instructional practices in reading and writing, and academic rigor (using the Instructional Quality Assessment). Student achievement measures of reading, writing, and mathematics will be drawn from standardized tests.

**Data Analytic Strategy:**
Using a group randomized design, researchers will conduct various fixed-effects multi-level analyses of the effect of the intervention on school practice. For these analyses
schools are the unit of random assignment; however, measures of the quality and the impact of *The Learning Walk* treatment on the development of school practice are taken at the level of the teacher. To evaluate the effect of *The Learning Walk* treatment on student achievement, researchers will also conduct several analyses that are variations of a basic random intercepts model with baseline covariates. In these analyses, schools are the unit of random assignment, with achievement outcomes nested at the individual student level. Researchers will also conduct qualitative analyses of interview and field note data in order to supplement the statistical analyses of the full random sample.
Learning Leadership: Kernel Routines for Instructional Improvement

RESEARCH NARRATIVE

1.0 Overview and Significance

Our randomized intervention study will test the efficacy of a strategy for developing school leaders that centers on developing leadership practice for literacy teaching through the structured and scaffolded implementation of a kernel organizational routine – The Learning Walk® routine. The Learning Walk® routine is a highly promising organizational routine for building a shared instructional vision, developing staff capability, and developing the school organization to support improvement in classroom teaching and student learning. We define The Learning Walk® routine as a kernel routine; that is, an organizational routine that has the potential for long-term transformation of school practice by seeding and propagating new forms of professional activity firmly tied to the school’s core work. The hallmarks of a kernel routine approach to school leadership development include:

a) Developing leadership practice by redesigning the school organization;
b) Developing leadership practice in situ – in actual schools and on the job;
c) Developing leadership practice via leadership teams in a school, not just the principal;
d) Developing leadership practice firmly anchored in classroom teaching and student learning activities.

We now motivate and justify our efficacy study of The Learning Walk® routine as a leadership development strategy and describe the approach in five steps, each organized into a separate sub-section. First, school leadership is a key mechanism for forging tighter coupling between (federal, state, and school district) policy directives and classroom teaching. Tighter coupling between policy and classroom teaching is essential if recent policy initiatives (e.g., NCLB) are to improve student learning (Section 1.1). Second, tighter coupling among policy, administration, and teaching in most American schools necessitates changing the practice status quo through re-designing organizational routines. Organizational routines structure and shape school practice and therefore offer a critical leverage point for transforming school practice (Section 1.2). Third, we offer the concept of kernel organizational routines as a powerful mechanism for reshaping school practice necessary for improvement in classroom teaching and student learning (Section 1.3). Fourth, re-envisioning school leadership as diagnosing, managing, and modifying organizational routines provides a new leverage point for leadership development (Section 1.4). Finally, we describe a particular kernel routine (The Learning Walk® routine) that will be the subject of our efficacy study (Section 1.5).

Section 1.6 lays out the logic model that guides our efficacy study. Although past research on leadership documents an association among leadership, school conditions, and the likelihood of improved student performance, it does not yet demonstrate a causal relationship. Further, there is no empirical evidence to support a causal relationship between particular leadership development programs and leadership practice. Our efficacy study is designed to demonstrate a causal relationship between a particular approach to leadership development that is grounded in teaching and learning, produces specific changes in school practice, and thereby creates measurable improvement in

---

1 The Learning Walk® (hereinafter referred to as “LW”) is a registered service mark of the University of Pittsburgh.
1.1 Coupling Policy and Teaching: The Importance of School Leadership

School leadership is critical to recent policy initiatives aimed at influencing classroom teaching and contributing to improved student learning. Over the last two decades, the movement toward standards and test-based accountability for achievement has created historically unprecedented external pressure on what is taught and—inevitably—how. U.S. education policy has moved away from the standard American pattern of local control over education and the resulting “weak coupling” among policy, school leadership, and classroom practice (Weick, 1976). Responding to the demands for better test performance, many school districts are creating systems of “managed instruction”: instructional guides that lay out a district-wide teaching program and interim “benchmark” tests that check whether students are on track toward year-end expectations. In this environment, principals are expected to be magicians of “instructional leadership”—a role that is largely undefined and for which few have any meaningful preparation.

These policy initiatives aim to establish tight coupling among policy, administration, and the technical core—classroom teaching and student learning (Rowan & Miskel, 1999; Rowan, 2002; 2007; Spillane & Burch, 2006), and represent a dramatic shift in the functioning of American schools. School leaders are under particular pressure in this changing policy environment as policymakers increasingly hold them accountable for improvement. Schools—not individual students or teachers—are the unit of measurement in both the national system (NCLB) and its state variants. Leaders have no door to close, behind which they can continue to do things the old way—managing bus, lunch, and athletic schedules; keeping parents satisfied; and generally leaving instructional decisions to teachers. Principals can lose their leadership positions as a result of low performance of students on accountability tests.

Various lines of inquiry, from the “effective schools” research (Bickel & Beaujean, 2005; Purkey & Smith, 1985) to research on the school as a professional community (Rosenholtz, 1989; Louis, Kruse & Bryk, 1995; McLaughlin & Talbert, 2006), have repeatedly noted the importance of the principal in enabling school improvement and high levels of student learning—especially among poor and minority students (Leithwood et al., 2007). Similarly, implementation scholarship has long identified the school principal as a key factor in the successful implementation of state and federal policy (Berman & McLaughlin, 1977). Principal leadership is important in promoting school-level conditions, i.e., shared vision for instruction and norms of trust, collaboration, and collective responsibility for students’ academic work that create incentives and opportunities for teachers to improve their practice (Bryk & Driscoll, 1985; Louis, Kruse & Bryk 1995; Rosenholtz, 1989). Syntheses of the literature suggest a positive relationship between the work of principals who develop and communicate school goals, monitor instruction, and mobilize incentives for teachers, and the work of teachers who are committed, open to innovation, and professionally involved (Shepard, 1998). The evidence also suggests that principal leadership, as mediated through the development of these school-level conditions, has an effect on student learning (Hallinger & Heck, 1996; Leithwood et al., 2007).

Three key organizational functions are consistently identified that school leaders must address for their schools to improve (Leithwood et al. 2007; Bryk & Schneider,
1996; Bryk & Driscoll, 1985; Newman & Wehlage, 1995); setting direction, developing the staff, and developing the organization. First, setting direction and maintaining a direction are essential (Bryk & Driscoll, 1985; Newman & Wehlage, 1995). School leaders must build a vision, develop goals to help staff reach that vision, and win buy-in from staff for that vision and goals. Second, developing the staff means that school leaders must build human capital by continuous monitoring of classroom teaching, student learning, and program implementation; supporting staff in further learning; and recognizing individual successes (Hallinger & Heck, 1996; Newman & Wehlage, 1995; Leithwood et al., 2007; Purkey & Smith, 1985; Sergiovanni, 1998). Third, essential to developing the organization, school leaders must create norms of trust, collaboration, and collective responsibility for student success; maintain an orderly work environment; and procure and distribute resources equitably (Bryk & Driscoll, 1985; Bryk & Schneider, 1996; Rosenholtz, 1989).

Alone, principals cannot fulfill these three sets of organizational functions (Heller & Firestone, 1995; Camburn, Rowan, & Taylor, 2004). Instead, it is necessary to take a distributed perspective on leadership and create schools in which lead teachers and instructional coaches play a significant role in the work (Spillane, 2006). Nevertheless, most leadership development programs focus on the school principal as an individual; they emphasize the knowledge and expertise of principals with the intention of changing how they practice on the job but—in part because they are often special institutes outside the school system—do not work on developing leadership practice in situ. Further, even when these programs include internships or mentoring support on the job, they rarely provide systematic processes and tools for meeting the core development functions that are recognized as key to school improvement. We propose to test a leadership development strategy, The Learning Walk® routine, that focuses on leadership teams rather than an individual, develops leadership practice in situ, and provides systematic processes and tools for school improvement. Our development strategy is consistent with the available empirical evidence on professional development in education. Specifically, activities are most effective at fostering teacher change when they involve collective participation of teachers from the same school, grade, or subject, are linked to teachers’ on-going daily activities, provide active learning opportunities, and are content-specific (Birman, Desimone, Porter, & Garet, 2000; Garet et al., 2001; Desimone, Porter, Birman, Garet, & Yoon, 2002).

1.2. School Practice and Organizational Routines

As with any organization, schools function through a set of interconnecting routines “repetitive, recognizable patterns of interdependent actions, involving multiple actors” (Feldman & Pentland, 2003, p. 311). These routines are critical for any organization to function more or less effectively because they provide stability and continuity over time (Feldman, 2000; Feldman & Pentland, 2003; March, 1981; March & Simon, 1958), and structure action in organizations (Allison, 1971; Gersick & Hackman, 1990; March & Simon, 1958; 1993). Theorists, March and Simon, argued that individuals cannot routinely use fully rational decision-making because of inherent limits in information processing capacity. Instead people satisfice—i.e., find a workable but not necessarily perfect solution rather than attempting to continually optimize. Organizations, they argued, do the same. Groups and individuals in the organization develop routines
that constitute the normal ways in which work gets done. These routines are not always in
the official manuals, but they allow members to perform satisfactorily, in the judgment of
clients and supervisors, and for their own self-satisfaction. Such routines often involve
adaptation to internal and external institutional constraints and may also recruit the power
of informal “below the radar” work groups, as documented by socio-cognitive research
(Orr, 1996; Suchman, 1996; Brown & Duguid, 2000; Resnick, Saljo, Pontecorvo, &
Burge, 1997).

Research has documented the ways in which organizational routines frame and
enable interactions, provide stability across time, and assist in socializing new
organizational members (Feldman & Pentland, 2003; Cohen & Bacdayan, 1994; Sherer
& Spillane, in press; Spillane, Mesler, Croegaert, & Sherer, 2007). Their very
pervasiveness and efficiency, however, together with the fact that they often function
without official or explicit recognition, can make routines also function as inhibitors of
innovation (Hannan & Freeman, 1984). People in organizations often resist disruption of
their ongoing practice—understandable in light of the great personal and group “costs” of
changing established practice (Marris, 1974). The more complex the organization, the
more stable the personnel, the more demanding the external demands—the more
members resist changes in routines.

1.3 Redesigning School Practice: “Kernel Routines” for Organizational Change

Just as existing routines work to stabilize organizations so, too, new routines can
serve as sources of change (Feldman & Pentland, 2003; Sherer & Spillane, in press;
Spillane, et al., 2007). When chosen purposefully and implemented well, new
organizational routines can function as powerful instruments for transforming school
practice. Resnick and Spillane (2006) used the term kernel routine to denote an
organizational routine that has the potential for transforming school practice by seeding
and propagating new forms of practice in schools. The idea is to introduce a routine
that—because it is highly specified and supported by well-defined tools and strategies—
can be implemented quickly at a reasonable level of quality under the guidance of the
principal. The routine has to be visibly focused on teaching and learning and responsive
to established standards of accountability in the school. Kernel routines serve two core
goals. First, a kernel routine anchors school practice in teaching and student learning.
Second, a kernel routine connects and weaves together other organizational routines in
the school to more fully focus the organization’s attention on instruction and learning,
rather than on institutional compliance.

For kernel routines to supplement existing routines or supplant less productive
ones, they must be sufficiently specified, developed, and scaffolded to change the way
people work. By sufficiently specified we mean clear articulation of the steps in the
routine, the rationale for these steps, and the requirements of each step. This calls for
training procedures and a set of tools and artifacts for performing the routine.

In our theory of kernel routines, leaders are not just enactors of a prepared script.
To function as a kernel for organizational change, the routine that the leader introduces
must be designed to encourage a process of appropriation (Wenger, 1998) in which users
adapt the routine to their particular conditions and capabilities. They do this by
developing new specific forms of the routine and new related routines over time. It is this
openness to, and encouragement of, appropriation that makes the routine a kernel for
organizational change. Hence, although kernel routines have to be well specified and developed to ensure high fidelity implementation at the outset, they must also enable appropriation and adaptation to seed and propagate new school practice.

In the first phase of implementation, kernel routines are introduced for faithful high fidelity implementation with their original design. Through training and scaffolded performance of the routine, school leaders, and then classroom teachers, learn to perform the routine in ways that are consistent with its designed intent. In the second phase, a release occurs from the performance of the specifics of the kernel routine that allows for the evolution of new routines and the re-design of existing ones in the school.

In order to seed and propagate work in schools, a kernel routine must meet the following six criteria.

- First, it must be centered on the technical core–teaching and student learning.
- Second, it must be anchored both in the official curriculum of the district and the enacted curriculum of the classroom–what is actually delivered to students.
- Third, it must build common understanding about teaching and learning among district and school staff members.
- Fourth, it must build trust and mutual access among school staff members.
- Fifth, it must provide routes by which new knowledge can enter the school’s community of practice.
- Sixth, it must be open to transformation over time without loss of its core designed elements.

1.4 Reconceptualizing School Leadership as the Redesign of Organizational Routines

Implementing and institutionalizing kernel routines in schools leads one to redefine and re-envision leadership practice as the management and guidance of routines that underlie the work of schools and that can lead to improvement in that work. A principal and school leadership team must understand the routines already at work in the school, and then systematically introduce selected new routines—kernels—that, over time, can replace ones that block change, resulting in new routines that promote more productive professional practice (Resnick & Spillane, 2006; Mabey & Iles, 1994; Spillane, et al., 2007). School leadership practice envisioned in this way fosters attention to direction setting, developing staff, and developing the organization. Careful selection of a kernel routine anchored in both the official curriculum and the enacted curriculum necessarily sets a direction for instruction. Moreover, by simultaneously attending to both the designed and the enacted curriculum, appropriately designed kernel routines can increase the chance that this direction setting establishes realistic and attainable goals for school staff. Kernel routines provide vehicles for sharing teacher knowledge within the building and for recognizing when that knowledge has been tapped out and new sources are needed to support the change process, hence developing staff capability. Finally, kernel routines can be designed to build trust, collaboration, and collective responsibility for student learning.

Leading by managing routines is consonant with recent scholarship on the distributed nature of school leadership (Spillane, 2006; Spillane & Diamond, 2007).
Scholars have increasingly argued for moving beyond equating school leadership exclusively with the principal’s office—this narrow view of leadership fails to capture the complexity of the work (Heller & Firestone, 1995; Ogawa & Bossert, 1995; Pitner, 1988; Katz & Kahn, 1966; Barnard, 1968; Heenan & Bennis, 1999) and fails to recognize that regular classroom teachers also play important roles in leading the school (Smylie & Denny, 1990; Heller & Firestone, 1995; Camburn, Rowan, & Taylor, 2004; Spillane & Diamond, 2007; Spillane, Camburn, & Stitzel, 2007). A recent study of more than 100 U.S. elementary schools found that responsibility for key organizational functions was typically distributed across three to seven formally designated leadership positions per school (Camburn, Rowan, & Taylor, 2004). Another study of 42 school principals using Experience Sampling Methodology (ESM) found that both formally designated leaders and individuals with no formal leadership designations took responsibility for leading and managing the school (Spillane, Camburn, & Stitzel, 2007). By preparing principals and leadership teams within schools to implement a kernel organizational routine, the approach to leadership we propose recognizes that school involves more than just the school principal. Moreover, it focuses on the practice of leadership rather than focusing exclusively on leadership positions, structures, and styles (Eccles & Nohria, 1992; Hallinger & Heck, 1996; Heifetz, 1994; Spillane, 2006; Tucker, 1981). Our study uses The Learning Walk® routine to transform leadership practice via training and scaffolding of the implementation and institutionalization of a kernel routine focused on literacy instruction and learning.

1.5. The Learning Walk® Routine as a Kernel Organizational Routine

Imagine a group of school staff visiting classrooms in their own school. Their classroom visits are part of the LW initiative that is designed to transform their school. A team composed of the principal, a coach, and three teachers enter a fourth grade classroom. This kind of visit is so routine that it evokes only a nod from the teacher. Students continue their work without interruption. A read-aloud of The Upside Down Boy by Juan Felipe Herrera is in progress. Students are discussing the main character in the book who is an immigrant and feels “upside down” because he doesn’t understand or speak English yet and is confused about school routines such as recess and behavior in the cafeteria. The students discuss the symbolism of borders in this book and identify borders or barriers in their own lives that they want to traverse. On the wall is a large chart listing four books by Herrera and across the top are listed schema categories for an author study: what the books are about, range of genres, elements of style and craft, and significance of the author in the world of literature. A visitor notices samples of student writing from another unit posted on the wall, with written feedback from the teacher, and a criteria chart for good writing posted next to the work. A second visitor concentrates on the teacher’s talk, trying to understand whether the teacher is reading this book to the class for the first time or whether this is a rereading with the intent of comparing it to other texts by this author. A third visitor examines students’ writing in their reader response journal. The last two visitors talk with students and ask questions such as, “What are you learning today?”, “What are you working on?”, or “How will you know if your work is good?” After ten minutes, the team moves to the hall where they briefly describe their observations and raise questions about the intent of some of the teachers’ actions. After a few minutes, they move into another classroom and repeat the process.
At the end of the day, the team meets with the teachers whose classrooms were observed. The team describes what they observed and the questions that emerged during hallway conversations. The classroom teachers make comments, take notes, and raise additional questions. The literacy coach wonders what might be heard from students if they were internalizing the schema for an author study. Teachers talk about the schema categories on the wall chart and ask if there was evidence in student journals of themes the author writes about or references to web resources about the author by others. One walker notes that several students in the classroom could name barriers they or their families have encountered similar to *The Upside Down Boy*. The coach presses walkers to articulate the question asked by the teacher to elicit this response. Hearing the exchange a teacher says, “I’m going to try that” and another teacher says he will, too. But he wonders how students will transfer what they learn from talking about these complex ideas to writing about them. A lively discussion follows and both teachers ask the coach to help them plan an arc of lessons with writing assignments on authors they are studying. The group then plans the date and focus of the next round of learning walks, which will occur in a couple of weeks, with three of the teachers who were observed this time being observers next time.

What you have just “witnessed” is a small sample of The Learning Walk® routine. This routine, and its associated training and observational tools, were developed by the Institute for Learning (IFL) at the University of Pittsburgh as the core of its leadership development program (Institute for Learning, 2004; Resnick & Glennan, 2002). As a kernel organizational routine it is comprised of a highly structured set of activities for the observation and interpretation of teaching and learning (Appendix B, Exhibit 1) based on the IFL’s Principles of Learning (Resnick & Hall, 2003). The LW routine and its accompanying training by IFL involve intensive practice-based learning for school leaders, comprised of a team of the principal and three additional leaders per school. The LW routine is meant to be implemented initially with the specific sequence of steps taught (see Figure 1, Column 2), but is also designed to generate new routines and transform existing ones in the school (see Figure 1, column 3).

The LW routine meets all six criteria for a kernel routine shown in the bulleted list on pages 4-5 (see Section 1.3). First, it is based on repeated observation and interpretation of teaching and student learning. Second, it is anchored both in the official curriculum of the school and the enacted curriculum of the classroom. Third, it uses research-based Principles of Learning (Resnick & Hall, 2003) and Content-Specific Classroom Observation Guides (Appendix B, Exhibit 3) to create a common language among participants. Fourth, the LW routine builds trust and mutual access among staff. Fifth, it provides routes by which new knowledge can enter a school’s community of practice through training, observations, and discussions. Sixth, it facilitates tailoring by school staff and is open to transformation over time.

The Learning Walk® leadership development system was initially inspired by leadership practice observed in New York City’s Community School District Two as part of the High Performance Learning Communities Project (Fink & Resnick, 2001; Elmore & Burney, 1999; Stein & D’Amico, 2002). In that district, a supervisory Walkthrough was used by the superintendent as a “high stakes” on-the-ground review of all elements of a principal’s instructional leadership activity. The IFL adapted Walkthroughs for use in its partner districts to focus participants on improving instruction and learning rather
than as a supervisory or evaluation tool. To mark this new and refined version of the routine, it was renamed The Learning Walk® routine.

As LW practice became widespread in IFL districts, the Institute codified and refined the practice through the development of a Suite of Tools, funded by The Wallace Foundation®, designed to promote faithful implementation and support going to scale in large districts. These tools include The Learning Walk® Sourcebook, an in-depth introduction to The Learning Walk® routine for teachers and administrators; The Learning Walk® Facilitator’s Guidebook (Appendix B, Exhibit 2) for leading LWs and using the tools for professional development; “The Learning Walk® Routine: A Tool For Getting Smarter about Teaching and Learning” (introductory video with accompanying slide presentation); The Learning Walk® Simulated Classroom Visits (six “portable” classrooms to be used for practicing the LW routine) and two extended The Learning Walk® videos. The IFL has worked with more than 50 districts and most of them continue to use The Learning Walk® routine.

Other “walkthrough” programs exist, but none is quite as elaborated nor has a training program equal to the IFL’s. A search of the professional literature confirms widespread use of semi-structured classroom visits in at least 17 states including those with the largest school enrollment (New York, California, Texas, and Florida) and 32 school districts (including 13 of the 20 largest districts) as well as in Canada, the United Kingdom, and Australia. The professional literature is replete with enthusiastic reports by educators of classroom visits becoming a way of life at their schools. There are reports from schools that have experienced gains in student learning and attribute those gains, at least in part, to LW-type practice (Abrutyn, 2006; Barnes & Miller, 2001; Keruskin, 2005). Schools attribute other improvements to LW-type practice, as well, including focus on student learning and achievement; focus on standards; greater familiarity with curriculum; collective understanding of instructional practice; more significant professional development resulting in improved instruction; improved teacher motivation and reflection; collegiality and trust among staff; and increased student engagement and decreased discipline problems (Abrutyn, 2006; Barnes & Miller, 2001; Blase & Blase, 1999; Ginsberg, 2001; Hopkins, 2005).

Despite this widespread professional enthusiasm, there are to our knowledge no formal evaluations or efficacy studies of this approach in the research literature. Three studies of IFL program implementations strongly suggest positive effects, however.

The most complete study of IFL leadership training, including LWs, was conducted by MDRC (Quint, Akey, Rappaport, & Willner, 2007). MDRC studied IFL leadership training practices in elementary schools in three school districts in different parts of the country. The study evaluated a logic model in which leadership training involving The Learning Walk® routine was expected to produce greater engagement of principals in instruction-focused professional development for their teachers. Principal engagement with teachers was expected to produce higher quality classroom instruction, which, in turn, was expected to result in improved student learning as measured by accountability tests used by the districts. The study found statistically significant relationships between each step in the theory of action and the next step. However, because all data were collected within a single school year, and there was no formal untreated comparison group, it was not possible to establish causal effects of The Learning Walk® training. It is this very causality that will be tested in our proposed
intervention study here.

A research team headed by Joan Talbert of Stanford University evaluated an IFL intervention in Austin, Texas, designed to build strong, instructionally focused professional learning communities in high schools, through training in discipline-focused instructional design accompanied by The Learning Walk® classroom visitation in six urban high schools. The evaluation report (Talbert & David, 2008) suggests that the IFL intervention provides an effective vehicle for developing teacher collaboration centered on instruction, as well as for increasing the academic rigor of teaching and learning. A similar high-school level intervention in Los Angeles, combining subject matter training and teacher leader development, including structured classroom observations, yielded similar results (David & Greene, 2007).

1.6 Logic Model for Research Design

Theory-driven evaluation involves using the substantive theories about the relationships between a program’s treatment variables and outcome variables to guide the design of the evaluation (Chen & Rossi, 1983; Shadish, Cook, & Leviton, 1991). We use a theory-driven evaluation approach because it facilitates strong causal inferences on efficacy, as well as enabling contributions to basic social theory (Birckmayer & Weiss, 2000; Chen & Rossi, 1980, 1983; Lipsey & Wilson, 1993). Our logic model (Figure 1) captures how our efficacy study is framed by the social science research. The model specifies that IFL’s training in The Learning Walk® routine (Column 1) will produce high fidelity implementation of the routine (Column 2). Implementation of the LW routine will in turn produce changes in school practice (Column 3), which will then produce changes in instruction and learning (Column 4).

The intervention to be assessed is IFL training of principals and their lead teachers in The Learning Walk® routine (see the Training and Implementation Plan, Appendix B, Exhibit 5). Principals, with their teams, will implement the eight steps of The Learning Walk® routine (see Appendix B, Exhibit 1) multiple times. At the outset, we expect high fidelity implementation, with school leaders following the steps as prescribed and directly using the tools provided.

Our logic model posits that over time, as school leaders implement The Learning Walk® routine with different subgroups of teachers, the routine that has been taught will be appropriated and modified and result in changing school practice (Figure 1, Column 3). Specifically, our model predicts changes in school practice on three dimensions of school practice—the norms of practice, the structure of practice, and the emergence of collaborative routines. The bi-directional arrows in the figure denote that these three dimensions interact.
Figure 1. The Learning Walk® Routine as Kernel Routine Theory of Action

With respect to norms, our logic model posits that the implementation of The Learning Walk® routine and its modifications will lead to improvements in norms of trust, collective responsibility for student learning, collaboration, and openness of innovation among school staff, all recognized aspects of strong professional communities. Our logic model is supported by theoretical and empirical work on innovation in schools and other organizations (e.g., Newmann, Marks, & Gamoran, 1996; Kruse, Louis, & Bryk, 1995; Talbert & McLaughlin, 1999). Researchers have examined variation in the degree to which teachers feel collectively responsible for student learning; have a shared commitment to high academic standards; trust their leader and one another; are open to innovation; and are reflective about their own practice. Many of these factors within a school have been correlated with higher teacher satisfaction (and retention), higher student engagement, student commitment to learning, and indeed higher student achievement (Bryk & Schneider, 1996; Newman & Whelage, 1995; Louis & Marks, 1998; Talbert & McLaughlin, 1999; Leanna & Pil, 2006).

Within this research tradition, researchers have developed strong theoretical bases for measures of school norms, but most of the empirical evidence has been correlational; researchers have been limited in their ability to draw causal claims (Newmann, Marks, & Gamoran, 1996). Little is known, therefore, about the extent to which deliberate interventions can be successful in developing norms of school practice and influencing student learning. Our study represents a significant departure from this tradition both...
because it is a randomized comparison study and because we will follow schools longitudinally enabling us to overcome threats due to selection effects.

With respect to **structure of practice**, our model posits that school staff will interact with *more colleagues* more *frequently* about instruction and student learning. *The strength* of ties among staff will increase over time. Further, the model posits that these interactions will increasingly *span grade levels* to include teachers in other grades and school leaders and individuals beyond the school.

Tie strength and tie span—two key aspects of structure—are important for innovation in organizations. Prior research suggests that strong ties are necessary for the transfer of tacit, complex, and sensitive knowledge (Uzzi, 1997; Reagans & McEvily, 2003), the sort of knowledge that is often critical for improving classroom teaching. Strong ties also support joint problem solving among organizational members (Uzzi, 1997). Recent work has examined the extent to which social capital is utilized to influence reform implementation (e.g., Frank, Zhao, & Borman, 2004) and how access to reform “expertise” within social networks influences teachers’ instructional practices (Penuel, Frank & Krause, 2006). With respect to tie span, interactions that span "multiple knowledge pools" (Reagans & McEvily, 2003, p. 242) reaching beyond their immediate grade level, or even school, allow school staff to access new information about instruction. One recent study of 88 urban schools, for example, concludes that a school’s internal and external ties, or social capital, predict student achievement (Leana & Pil, 2006).

With respect to **collaborative routines**, our logic model posits that the implementation of the LW routine will result in more school staff engaging in more collaborative instructional planning, studying student work, analyzing data, and critiquing lessons. These collaborative routines are expected to contribute to teachers’ knowledge base, professionalism, and ability to act on what they learn. When teachers work together to explore concrete connections between practice and outcomes, they create a setting in which discussion and reflection of data results in new understanding and motivation for change (McLaughlin & Talbert, 2006). Through practice and exposure to expert personnel, leaders and teachers will improve in their ability to critique lessons becoming more “expert” in various capabilities associated with effective teaching.

The final link in our logic model specifies that positive shifts in the norms and structure of school practice and in collaborative routines will lead to improvements in instruction and student learning (Figure 1, Column 4). It has been well established that students’ opportunities-to-learn (OTL) content are important for student learning (see, for example, Schmidt and Maier, 2009; Cooley and Leinhardt, 1980; Gamoran, Porter, Smithson and White, 1997; Rosenshine and Stevens, 1986). Additionally, studies have shown that changes in instructional content are sensitive to interventions in schools when supports for teacher learning are supplied (e.g., Correnti and Rowan, 2007). Given the focus of The Learning Walk® routine on identifying academic rigor we would especially expect increase in the frequency with which high cognitive demand content is taught (Correnti and Phelps, in preparation). We will measure student learning by changes in results on statewide standardize tests. We will track changes in instruction by surveying teachers in treated and comparison schools about their instructional practices in reading and writing, collecting and analyzing samples of teacher task assignments and student work, and through observations of focal teachers in the eight case study schools (half
2.0 RESEARCH METHODOLOGY

2.1 Study Design

We will study the effects of the IFL Learning Walk® training in a group-randomized comparison trial. Randomization will occur at the school level to prevent to the extent possible, spillover effects to the comparison condition. Random assignment will permit us to draw causal inferences about the influence of the LW training on school practice, conceptualized in terms of collaborative routines, the structure of practice and norms of practice. These changes in school practice will influence the quality of classroom teaching, which will ultimately be reflected in greater student achievement gains in the treatment than in the comparison group. We will use a school staff survey in treatment and comparison schools, student achievement data, and samples of teacher and student work. In addition, we will conduct a set of embedded case studies to examine how school-level conditions mediate the observed effects (or lack thereof) so that we can better understand how the intervention enables change.

Our research questions are:

**Research Question 1:** Does the IFL training generate greater participation staff in collaborative routines and produce the anticipated changes in the structure and norms of school practice?

**Research Question 2:** Does participation in the IFL LW training produce improvements in instruction and learning as measured by student achievement?

To address Research Question 1, we will use a school staff survey in both treatment and comparison schools to measure school practice conceptualized in terms of collaborative routines, structure of practice, and norms of practice both during the LW training period (first two years of the study) and for a full school year after intensive training has ended. This survey will also include items designed to measure the quality of classroom instruction. In addition, to the biannual survey items about teachers’ classroom practice, we will collect teachers’ assignments and student work samples. As part of our embedded case studies, we will also observe classroom instruction for a subset of teachers in four treated and four comparison schools. In treatment schools we will also use records of participation in the IFL LW training and of participation in the LW routine to measure treatment dosage and fidelity of LW implementation (see Section 2.4 below for details on these measures).

To address Research Question 2, we will measure student learning in order to conduct two types of analyses. From the student achievement point of view, we have an “intention to treat” or encouragement study (Gennetian, Bos, & Morris, 2002), in which the IFL LW training intervention encourages changes in school practice but does not directly manipulate those practices. We will begin by fitting models in an intent-to-treat analysis to understand if student learning is different in the treatment versus the comparison schools. We hypothesize that participation in the IFL LW training will produce differences between treatment and comparison schools in student gains on standardized achievement test scores in literacy. We will also examine the extent to
which growth in student learning is influenced by: a) participation in the IFL LW training, and b) the level of LW routine implementation at the school (see Section 2.4 below for details).

The formal quantitative analyses of the full randomized trial will be augmented and deepened by detailed embedded case studies in a sub-sample of schools. We will use an embedded case study approach to collect and analyze interview and observation data in eight schools, purposefully sampling schools from both the treatment and comparison groups. Combined, the evidence will provide insights into how the introduction of the LW kernel routine influences school practice (defined in terms of collaborative routines, structure of practice, and norms of practice) and student learning.

We provide a detailed discussion of our study and research methodology below. We begin by describing our sample of schools (Section 2.2) and then describe the intervention—the IFL LW training (Section 2.3). Next we describe our quantitative data collection efforts (Section 2.4) and our statistical models (Section 2.5). We then describe our embedded case studies (Section 2.6). Next we describe our power analysis and sample size determination (Section 2.7), cost-benefit analysis (Section 2.8) and conclude with a brief description of the study timeline and products (Section 2.9).

2.2 Sample Description
The study will be conducted in a large urban school district. Eighty schools (grades k-6) will be chosen from the district’s elementary schools and randomized into 40 treatment and 40 comparison schools. Reasons for selecting 80 schools are detailed in Section 2.6 where we conduct our power analysis calculations. Because we are especially interested in the extent to which the IFL LW intervention can aid at-risk student populations, we will use a group randomized design. We will begin by recruiting in the poorest elementary schools in the district for the study, by providing them with a comprehensive presentation about The Learning Walk® routine and our study. Specifically, we will focus elementary schools where greater than 50% of the students receive free or reduced lunch. After successfully recruiting schools we will then randomly assign schools to the treatment or comparison condition.

2.3 The Intervention: The Learning Walk® Training
School leadership teams in 40 elementary schools in the sample district, randomly selected from the district’s elementary schools, will be trained in the LW kernel routine by IFL and supported over a two-year period in its implementation. IFL will conduct an introductory session for schools that are considering participation in the study. For those schools that are randomly assigned to the treated group, LW training will follow a well-developed process for which training materials, training protocols, and a set of integrated tools for school use already exist. (These tools are described in Section 1.5 above).

Following baseline data collection for both Treatment and Control schools in September, the LW intervention will begin in October 2010. School leadership teams (the principal and three instructional leaders at each school) will undergo intensive training beginning October 2010. These teams will receive five full-day training sessions and five half-day training sessions. Principals will receive an additional three full-day and five half-day sessions. The content of the training sessions is outlined in Appendix B, Exhibit 5. This schedule reflects the importance of creating expert knowledge and buy-in from
the principal and, at the same time, emphasizes a distributed approach to leadership development by also focusing on the expertise within each leadership team.

The training schedule is based on prior experience working with districts, allowing for the gradual and successful introduction of the LW process in schools. From February 2011 forward, leadership teams will be expected to enact The Learning Walk® routine at least weekly in addition to participating in the IFL LW training sessions. The IFL training sessions are dispersed over two school years in order to capitalize on questions that arise when teams enact The Learning Walk® routine in their own schools and also to motivate teams to continue to enact the routine regularly so they become institutionalized. Teams of leaders and teachers will conduct a minimum of 12 learning walks between February and May (2011) in treatment schools. These walks will include at least one member of the leadership team and will incorporate other teachers as walkers until every teacher has been introduced to the routine. Further, they will rotate classroom visits until every teacher has been visited once. During the second year of the study, the leadership teams continue to be responsible for enacting The Learning Walk® routine, rotating teacher participation to provide ample opportunities for all teachers to visit classrooms and be visited by others. Learning walks will occur with a frequency of at least one per week. Although formal training by IFL staff concludes at the end of the 2012 school year, the intervention is intended to be self-sustaining. The continued collection of LW artifacts (e.g., standardized forms that are part of the LW system, including Focusing, Classroom Visit, Observation, Debrief, Next Steps sheets and LW logs as displayed in Appendix B, Exhibit 4), even after the training is concluded, will encourage adherence to its practice.

2.4 Data Collection
2.4.1 Data Collection for Research Question #1: The Learning Walk® Training (LWT) → Changed School Practice (SP)

General measures for both treatment and comparison schools. In order to collect the necessary data to test these hypotheses, we will use a combination of a bi-annual school staff questionnaire along with IFL and school records of participation in the LW routine. The School Staff Questionnaire (SSQ) will be web-based and given in both the treatment and comparison schools. The Institute for Social Research (ISR) at the University of Michigan will administer the SSQ and provide follow-up to produce response rates targeted at greater than 70 percent. Respondents will be paid an honorarium for completing the SSQ.

We will administer the SSQ in treatment and comparison schools beginning in late September of 2010, prior to the beginning of the treatment. As the IFL LW training does not begin until October 2010, the initial administration will serve as baseline data for our study. The SSQ will be administered twice per year in years 1, 2, and 3 of the study.

We will use the SSQ to collect data on school practice–collaborative routines, structure of practice, norms of practice, and classroom instruction (i.e., proximal outcomes). As part of the SSQ, we will also collect data on leadership team members’ expertise in judging and critiquing the quality of classroom instruction. These measures will be taken on leadership teams in both treatment schools and comparison schools. Because the school district will withhold training in the LW kernel routine from
comparison schools for the entire three-year period of the study, we will have a multiple-
group design that will allow us to compare treated and comparison schools
longitudinally. This multiple group design will allow us to rule out history and
maturation as threats to internal validity.

To collect data on treatment dosage (Boruch, 1997; Rossi, Lipsey and Freeman,
2004), district and IFL staff will keep detailed logs of participation in training sessions
over the course of the intervention. We will also collect data to track the level of
implementation of the LW kernel routine in treatment schools over the three years of the
study. We will do this in two ways: First, we will collect documents (artifacts) from
learning walks conducted in the treatment schools. This type of data collection method
minimizes errors associated with recall bias. Second, we will use the SSQ to collect data
on school staff participation in the LW activities. As the SSQ will be given in both
treatment and comparison schools, these data will enable us to understand if these
activities also occur in any of the comparison schools.

Specific measures of school practice in the treatment and comparison schools. To
understand whether The Learning Walk® training produces changes in school practice,
we will examine outcomes that measure collaborative routines, structure of practice,
and classroom instruction and compare mean effects between treated and comparison schools. These constructs were chosen because they measure
characteristics of The Learning Walk® intervention that are theoretically important for
school improvement design – that is, they measure the extent to which leadership practice
is distributed among leaders in the school, whether the practice is developed \textit{in situ} and
results in changes in the organization, and whether leadership practice is anchored in
classroom practice and student learning.

In examining whether \textit{collaborative routines} are established in treated schools we
will measure whether collaborations are occurring over teaching. Many of these measures
will be adapted from existing teacher surveys (see Table A1 in the appendix, where we
also list the source of the measures and their reliability in prior studies). We have chosen
items that ask specifically about instructional and assessment practice – i.e., the enacted
curriculum, teaching methods used by teachers, standards for student learning,
assessment practices used to demonstrate student learning, and the alignment of the
curriculum with teaching and assessment practices.

We will also measure leadership team members’ expertise in judging the quality
of and critiquing classroom instruction. Through interactions with expert personnel,
leaders and teachers improve their ability to critique lessons becoming more “expert” in
various capabilities associated with effective teaching. A long tradition of cognitive
research contrasting “experts” and “novices” in various fields (deGroot, 1946/1965;
Chase & Simon, 1973; Chi, Glaser, & Farr, 1988; Cuthbert, et al., 1999) show that
experts in a field make faster, more complex, and more accurate judgments of situations
they encounter. To measure team members’ expertise we will use scenario simulations
(incorporated into the SSQ) as item prompts (following Kersting, 2008). This approach to
measuring instructional knowledge expertise has been used by education scholars
including Sabers, Cushing & Berliner, 1991; Tsui, 2003; and Spillane, White & Stephan,
2007. Study participants will watch brief video clips of teaching and review samples of
student work and write about what they observe and what if any questions/feedback they
might have. Participants’ responses to these simulations will be scored on a rubric based
on the Principles of Learning. In previous work, researchers have demonstrated adequate inter-rater reliability (Crosson, et al., 2006; Junker et al., 2006) and predictive validity (Junker, et al., 2006) of a similar instrument examining instructional quality.

With respect to **structure of practice**, two measures will be central – tie strength and tie span (see Section 1.6). To calculate **tie strength** we will use a combination of frequency of interactions (e.g., daily, weekly, etc.), and the level of influence a school staff member reports for an interaction with a particular individual. To calculate **tie span**, we calculate the average number of ties school staff have that extend beyond their immediate grade level and the school. With respect to substance of interactions, we will track shifts in tie strength and tie span in both treatment and comparison schools by school subject and examine which dimensions of instruction (e.g., content coverage, teaching strategies) teachers report talking about in their reported ties.

With respect to **norms of practice** we will examine whether norms in treatment schools develop differently from norms in comparison schools for such constructs as collective responsibility, collaboration, trust, and innovation (See Table A1 in Appendix for sample measures and items). We will examine these measures longitudinally in treatment and comparison schools to understand whether they develop differently over time. These constructs were chosen to examine the extent to which leadership practice is successful in developing the organization. Once again, items assessing these constructs will be grounded in specific contexts of instructional and assessment practices of teachers.

We will track changes in classroom instruction in two ways. First, we will survey all teachers in treated and comparison schools about their instructional practices in both reading and writing. These items will be used to understand the extent to which teachers in the different conditions focus on different types of literacy content and to what extent the instructional opportunities they provide to students changes over time as a result of learning walks. Second, we will collect teacher assignments and student work in order to assess the academic rigor of student-teacher interactions over content. We will measure academic rigor through the Instructional Quality Assessment (IQA) – a system of collecting teacher task assignments and student work. Task assignments are rated on several dimensions including their overall cognitive demand, the quality of the text being used, and teacher expectations for high quality student work. The IQA has been used in several studies that have shown differences between IFL-trained teachers and those not receiving this training (Matsumura, Garnier, Junker, Resnick and Bickel, 2009) and has also been related to student achievement (Matsumura, Garnier, Pascal, and Valdes, 2002; Matsumura, Slater and Crosson, 2008).

In addition to these quantitative measures of classroom practice, our embedded case studies will allow us to conduct in-depth qualitative observation and description of focal teachers in the eight case study schools (half treatment and half control schools). These data will provide a deeper look at differences in instruction among schools that experienced systematic LW training and those that did not.

### 2.4.2 Data Collection for Research Question #2: The Learning Walk® Training (LWT) → Achievement

Data collection for the second research question will be based on results of standardized testing in the district schools. For example, we would work with the district
to obtain test vertically equated scale scores for individual students as well as student reading level scores. These scale scores will allow us to examine individual students’ growth trajectories over time. Additionally, we will work with district and/or state personnel to obtain individual student-level data as well as identifying information that would indicate each student’s classroom teacher in a given year from Spring 2010 through Spring 2014. We will use these data in multi-level models predicting student achievement (see Section 2.5.2 below). While the intervention centers on literacy, we will also analyze achievement in mathematics to examine whether the intervention transfers to mathematics achievement.

2.5 Statistical Models and Other Analyses

2.5.1 Multi-Level Models for Research Question #1: The Learning Walk® Training (LWT) → School Practice (SP)

In our group-randomized design, schools are the unit of random assignment, and measures of the quality and impact of LWT on the development of school practice (SP) are taken at the level of teachers. School practice in these models refers to each of our separate measures of collaborative routines, structure of practice, norms of practice and teachers’ classroom practice as described in Section 2.4.1. This design is most naturally analyzed using elaborations of a two-level random intercept model (e.g., Snijders & Bosker, 1999, p. 63.):

Level 1 (Teachers): \((SP)_{jk} = \beta_{0k} + R_{jk}\)

Level 2 (Schools): \(\beta_{0k} = \gamma_{00} + \gamma_{10}(LWT)_{k} + \gamma_{20}(Prior SP)_{k} + U_{0k}\)

where \(R_{jk}\) and \(U_{0k}\) are independent normal residual errors, and where

\((SP)_{jk}\) = Quality of School Practice (SP) measured for teacher \(j\) in school \(k\)

\(\beta_{0k}\) = Nonintervention quality of SP in each school

\(\gamma_{00}\) = Population intercept for quality of SP

\((LWT)_{k}\) = Condition of school \(k\) (LWT= 1, comparison= 0)

\(\gamma_{10}\) = Effect of LWT on quality of SP

\((Prior SP)_{k}\) = Baseline/pre-treatment quality of SP

\(\gamma_{20}\) = Population effect of \((Prior SP)_{k}\) on \((SP)_{jk}\)

The focus of our interest is \(\gamma_{10}\), the effect of the experimental condition (LWT) on SP, as measured by individual teachers’ actions and perceptions from the SSQ. Our model treats \(\gamma_{10}\) as a fixed effect (see Murray, 2001). The pre-treatment measure of SP at the school level, \((Prior SP)_{k}\), is included in the model since it is well known (Raudenbush, 1997; Bloom, 2004) that this can greatly improve the power of a multi-level analysis to detect a program (intervention) effect. We will also explore prior SP measures as baseline covariates at the teacher level to make the best tradeoff between lost degrees of freedom (df.) and improved precision of \(\gamma_{10}\). Thus, a random intercepts model with baseline covariates will be the basic model we consider in evaluating the effect of the LWT intervention on SP.

Several variations of the model will also be examined as follow-ups to our main analysis, including:

- **Heterogeneity of the intervention effect.** Several variables that are not fully
controlled in our study will be adjusted for in the analysis, including school average SES (free/reduced lunch) status and ESL status since our intervention is at the school level. We will also explore whether the LWT effect is heterogeneous across schools (e.g., by testing for a school random effect corresponding to \( \gamma_{10} \)).

- **Time effects.** Recall that schools in both the LWT and comparison groups will be followed for three years. We hypothesize that the effect of the LWT intervention on SP will accumulate over time; and so we will also explore models in which time has been added as a main fixed effect, but allowing differential growth curves in SP within treated and comparison schools (Murray, 2001). Variations on this model will allow us to consider nonlinear or heterogeneous growth curves, sub linear growth curves, and variation in growth curves at the school level rather than the teacher level.

- **Multivariate outcomes.** Since we will have several ways to measure teachers’ changing behaviors and attitudes as a result of LWT, we will also consider multilevel models for multivariate teacher-level outcomes. Such models can be approximated using conventional MANOVA machinery, adjusting for the multilevel structure nesting \( m \) teachers with intraclass correlation \( \rho_{ICC} \) within each school with the design effect term \( DEFT = \sqrt{1 + (m-1)\rho_{ICC}} \), but a more accurate and flexible alternative that we will also pursue is to build multivariate responses into the multi-level structure above, using general-purpose software for constructing and estimating hierarchically structured models, such as BUGS (Spiegelhalter, Thomas, Best & Lunn, 2003).

- **Single-Level Models for School Practice Outcomes.** While some measures of SP in each school are focused on individual teachers’ behaviors, attitudes and perceptions, others are focused on SP in a school. In this case the unit of randomization, school, is the same as the unit of analysis (SP within school), and simpler, single-level, one-way ANCOVA models such as

\[
(SP)_k = \beta_0 + \beta_1(LWT)_k + \beta_2(Prior\ SP)_k + R_k
\]

can be employed. In addition to this basic model, the same extensions will be explored, e.g. heterogeneity effects will be explored by including additional covariates in the model; accumulation of effect over time will be explored with longitudinal models, and multiple outcomes will be explored with the analogous MANCOVA models.

### 2.5.2 Multi-Level Models for Research Question #2: The Learning Walk® Training (LWT) \( \rightarrow \) Achievement

In our group-randomized design, schools (groups of teachers/classrooms) are the unit of random assignment, but achievement outcomes are measured at the individual student level (nested within teachers). Certain covariates are measured at the teacher or classroom level. This design is most naturally analyzed with elaborations of a three-level random-intercept model (e.g., Snijders & Bosker, 1999, pp63ff.):

- **Level 1 (Students):** \((Achieve)_{ijk} = \beta_{0jk} + R_{ijk}\)
- **Level 2 (Teachers):** \(\beta_{0jk} = \gamma_{00k} + U_{0jk}\)
- **Level 3 (Schools):** \(\gamma_{00k} = \delta_{000} + \delta_{100}(LWT)_k + \delta_{200}(Prior\ Achieve)_k + V_{00k}\)
where \( R_{ijk}, U_{0jk} \) and \( V_{00k} \) are independent normal residual errors, and where

\[
(Achieve)_{ijk} = \text{Achievement of student } i \text{ under teacher } j \text{ in school } k
\]

\[
\beta_{0jk} = \text{Nonintervention achievement level in each teacher's classroom}
\]

\[
\gamma_{00k} = \text{School-level intercept for achievement}
\]

\[
\delta_{000} = \text{Population intercept for achievement}
\]

\[
(LWT)_k = \text{Condition of school } k \text{ (LWT intervention= 1, comparison= 0)}
\]

\[
\delta_{100} = \text{Effect of intervention on achievement level}
\]

\[
\delta_{200} = \text{Population effect of } (Prior \text{ Achieve})_k \text{ on } (Achieve)_{jk}
\]

The focus of our interest is \( \delta_{100} \), the effect of the experimental condition (LWT) on student achievement. Our model treats \( \delta_{100} \) as a fixed effect (see Murray, 2001). We again include \( (Prior \text{ Achieve})_k \) to improve the power of the analysis (Raudenbush, 1997; Bloom, 2004).

We will also explore prior achievement as a baseline covariate at the teacher and student level to make the best trade-off between lost df. and improved precision in estimating \( \delta_{100} \). Thus, a random intercepts model with baseline covariates will be the basic model we consider in evaluating the effect of the LWT intervention on achievement. Several variations of the model will also be examined as follow-ups to our main analysis, including:

- **SP as a mediating variable.** Although SP is not manipulated directly in our study, an approach to studying its causal relationship with student achievement is available via “instrumental variables” (IV) estimation (Angrist, Imbens & Rubin, 1996; Gennetian, Bos & Morris, 2002). A variation on two-stage least squares (TSLS) analysis can be adapted to this setting (Spencer & Felding, 2000): We fit the two-level model for SP as indicated in Section 2.5.1 (Multilevel model for research question #1), and use predicted values of \( (SP)_jk \), from that model as the mediating variable in a three-level model for student achievement,

\[
\begin{align*}
\text{Level 1 (Students):} & \quad (Achieve)_{ijk} = \beta_{0jk} + R_{ijk} \\
\text{Level 2 (Teachers):} & \quad \beta_{0jk} = \gamma_{00k} + U_{0jk} \\
\text{Level 3 (Schools):} & \quad \gamma_{00k} = \delta_{000} + \delta_{100}(SP)_{jk} + V_{00k}
\end{align*}
\]

Effectively, \( (SP)_{jk} \) is the part of SP that is manipulated by LWT (and is therefore unaffected by any confounding variable that might otherwise influence both SP and student achievement, such as prior teacher quality), so that the estimate of \( \delta_{100} \) in this system of equations can be treated as a measure of the causal effect of SP on student achievement. Simultaneous estimation of these two multi-level models for instrumental variable estimation will also be explored.

- **Teacher- vs. School-level mediating variables.** Versions of the above analyses can be carried out using teacher-level measures of SP as mediating variables, or for school-level mediating variables. Both will be explored in looking for the best
account of the data we have collected.

- **Heterogeneity of the intervention effect.** Several variables that are not fully controlled in our study will be adjusted for in the analysis, including SES (free/reduced lunch) status and ELL status. Since our intervention is at the school level, we expect to include these variables at the school level as well. We will also explore whether the LWT effect is heterogeneous across teachers (e.g., by testing for a teacher random effect corresponding to $\delta_{100}$).

- **Time effects.** We hypothesize that the effect of the LWT intervention on student achievement will accumulate over time. We plan two analyses to test this hypothesis. First, we will explore a random coefficients/growth curve model focused on teachers, treating students as randomly equivalent from one year to the next. We will consider a model that retains a fixed main effect for the intervention, but allows differential growth curves in achievement for each teacher in each experimental condition (Murray, 2001). Variations on this model allow us to consider nonlinear, heterogeneous or sub linear growth curves, variation in growth curves at the school level rather than the teacher level, baseline covariates, and adjustments for student level covariates such as SES and ELL status. Differential growth in student achievement over time due to the LWT intervention is captured in $\delta_{100}$. Care will be needed in drawing inferences from this model due to selection effects as students move into and out of intervention (LWT) schools and grades from year to year. Alternative models to account for students passing in and out of intervention teachers’ classrooms, such as across-classified random effects model (see, e.g., Rowan, Correnti and Miller, 2002) will be applied to the extent that the data allow.

### 2.6 Embedded Case Studies.

As noted above, we will also conduct a set of embedded case studies to examine how school-level conditions mediate the observed effects (or lack thereof) on school practice, classroom teaching, and student learning to better understand how the intervention enables change; this will deepen our understanding of results related to Research Question #1. A case study approach is relevant for understanding complex processes such as organizational change (Erickson, 1986; Peshkin, 1993). Using a “constant comparative” approach we will collect data on comparable dimensions of school practice in the eight selected schools (Glaser & Strauss, 1967). Semi-structured interview and observation protocols with standard questions, that also allow for probing of and follow-up to respondents’ answers, will ensure comparable data is collected across all sites while simultaneously allowing for the unique perspectives of study participants to emerge.

We will observe and interview in a sub-sample of eight of the 80 schools in the randomized trial—four comparison and four treatment schools (10% of the treatment schools). Based on our analysis of data generated by the SSQ, we will select schools that differ on dimensions critical to the implementation and institutionalization of the LW kernel routine. Within each school, we will use a combination of random and purposeful sampling to identify eight to ten staff members (roughly 20% of the professional staff) to
interview each year. For example, we will construct a purposeful sample of individuals who are formal and informal leaders. In case study schools we will interview all four school leaders trained by IFL and two informal leaders; that is, teachers with no formal leadership designation but who are identified by colleagues as influential leaders. The SSQ will enable us to identify all the informal leaders in a school and we will randomly sample two from this group in each school. In addition, we will select four teachers (who are not informal leaders based on the SSQ) in each school.

We will devote the equivalent of five days to collecting data in each case study school, allocating approximately 60% of this time to structured observations (e.g., performance of LW and collaborative routines). We anticipate completing 20 hours of observations per school per year. This data collection schedule will allow us to observe multiple The Learning Walk® performances as well as numerous other collaborative routines.

Data collection and data analysis in the eight schools selected for case study work will be closely integrated with the ongoing quantitative analysis, allowing us to more deeply examine working hypotheses as they emerge, and refine qualitative data collection strategies as our work progresses (Miles & Huberman, 1984). A commercial data-coding program (e.g., HyperResearch) will be used to code and analyze interview and field note data, using both open and closed coding. We will use open coding on a randomly selected sample of our qualitative data to identify emerging themes that we will then use to define a set of coding categories that will be applied to the entire qualitative data base. We will also use our logic model to define coding categories in the qualitative data. Our qualitative analyses will supplement our statistical analysis of the full random sample, providing data for theory development and hypotheses building about causal mechanisms (thereby enabling us to go beyond establishing statistically that the intervention has an effect). The data generated from these embedded studies with the quantitative data will enable us to adopt mixed-method approaches (Caracelli & Greene, 1993).

2.7 Power and Sample Size Calculations

We plan to recruit a fixed number, $K$, of schools in the district that meet our intake criteria as described in Section 2.2. Half the schools will be assigned to treatment (LWT) and half to a comparison “standard practice” condition. In this Section we calculate that $K=80$ schools should be recruited, so that 40 will be randomly assigned to treatment and 40 to the comparison condition. The primary analyses outlined in the previous section reveal that the statistical analyses will be carried out with:

- A two-level model for the effect of LWT on univariate teacher-level measures of SP
- A one-level ANCOVA model for the effect of LWT on univariate school-level measures of SP
- A three-level model for the effect of LWT on a univariate student achievement measure

We examined power and sample size requirements for our study by first referring to effect sizes (differences between intervention and comparison conditions, divided by the standard deviation of the outcome variable) in the literature corresponding to our two-level models for SP and our three-level models for student achievement. Lipsey &
Wilson (1993) conducted a meta-analysis of 302 meta-analysis articles in education and psychology and found that the mean effect size for all studies was 0.45. Studies in the analysis specifically involving teacher training and effects on teacher and student outcomes had a mean effect size of 0.58. A meta-analysis of 91 research studies on in-service teacher education found, in general, a mean effect size of 0.52 (Wade, 1985). When data were grouped by the level at which the evaluation was directed, attempts to change participants’ behavior yielded a mean effect size of 0.60 and attempts to demonstrate results by looking at student learning had a mean effect size of 0.37. For our power analysis we conservatively assume effect sizes equal to or lower than those found in this literature search, as follows:

- The treatment effect for teacher (behavior and perception) outcomes is $\delta_T = 0.40$;
- The treatment effect for school level SP outcomes is $\delta_P = 0.60$;
- The treatment effect for student (achievement) outcomes is $\delta_S = 0.20$.

The intraclass correlation (ICC) measures the amount of within-group dependence between sampled units in clustered or nested data, such as is used with multi-level models. It is the ratio of the variance between groups to the sum of the within- and between-groups variances. For students’ reading achievement, Rowan, Correnti, & Miller (2002) found ICC’s for students grouped within teachers ranging from 0.12 to 0.23. King et al. (2002) found ICC’s for patient outcomes within doctors, in a study of a training intervention for doctors, ranging from 0.00 to 0.12. For SAT verbal scores, Everson & Millsap (2004) found ICC’s for students within schools of 0.15. Fletcher et al. (2004) computed ICC’s for a variety of self-report items on an inventory of science classroom practices. For teachers within schools they found ICC’s ranging from 0.12 to 0.26. For students within schools they found ICC’s ranging from 0.14 to 0.26. For our power analysis, we conservatively assume an ICC at the teacher level larger than those found in this literature search, as follows: an ICC between teachers within schools $\rho = 0.30$.

Finally, we assume that our pre-treatment (baseline) covariate(s)$^2$ will correlate approximately 0.7 with school-to-school variation within each study condition. Adjusting the model for such covariates would have the effect of reducing the teacher level ICC by about half, to approximately 0.15, which is conservative relative to the reductions reported by Rowan, Correnti and Miller (2002), and similar to those reported in Bloom (2004, Table 4.6). We used the program Optimal Design (Raudenbush, et al., 2004) to examine sample sizes needed for two-level models. We expect that:

- $J = 20$ teachers in each LWT or comparison school can be recruited to participate in the study;
- $n = 30$ students per teacher will be assessed, for all participating teachers.

Two level model for effect of LWT on teacher-level SP outcomes.

Figure A.1 in Appendix A indicates the power to detect a teacher level effect of size $\delta_T = 0.40$ at level $\alpha = 0.05$ with $J = 20$ teachers per school. The number of schools $K$ indicated along the horizontal axis would be divided evenly between LWT and comparison groups; the power is indicated along the vertical axis. The solid blue line in Figure A.1 shows that

---

$^2$ Baseline SP score, or baseline student achievement, depending on the outcome variable.
the design will have power $1 - \beta = 0.80$ for approximately $K = 80$ schools, if no baseline covariate is available. The dashed blue line shows that with a baseline covariate that explains approximately half of the pre-treatment variability in student achievement among schools, only $K = 45$ schools would be needed. If, for some reason, only $J = 10$ teachers per school could be recruited instead of $J = 20$, it would have the effect of increasing the school sample size by about 10 (increasing our necessary sample size from 45 to 55 schools).

**Linear model for effect of LWT on school-level SP outcomes.**
The analysis for school-level outcomes only involves an ANCOVA or OLS regression model. The same considerations as in the previous case applies; but now there is no hierarchical structure. The power analysis for this model was conducted using the program G*Power (Erdfelder, Faul & Buchner, 1996). In order to detect a school-level effect of $\delta_P = 0.60$, far fewer schools would be required compared to the above analysis for finding a teacher-level effect, or for the below analysis for finding a student-level effect. Indeed, in order to have power $1 - \beta = 0.80$ to detect an effect size of $\delta_P = 0.25$ it would require as few as 34 schools after adjusting for other covariates added to an ordinary least squares regression.

**Three level model for effect of LWT on student achievement.**
The analysis for student level outcomes is similar to the model for teacher-level outcomes, but now the model is a three-level model and we hope to detect an effect size of $\delta_S = 0.20$. Figure A.2 in Appendix A indicates the power to detect a student level effect at level $\alpha = 0.05$ with $J = 20$ teachers per school and $n = 30$ students per teacher. From Figure A2 we see that 130 schools would be needed if no baseline covariate is available. With a baseline covariate that explains approximately half of the pre-treatment variability in student achievement among schools, only $K = 70$ schools would be needed. If, for some reason, only $J = 10$ teachers per school could be recruited instead of $J = 20$, it would have almost no effect on the power to detect student-level effects.

**Final sample size determination; accounting for attrition.**
The above analysis shows that our sample size needs are essentially driven by the three-level model for student achievement outcomes. With a strong baseline covariate (e.g., last year’s test scores) a sample of $K = 70$ schools is needed, divided evenly between LWT (35 schools) and comparison (35 schools). We will recruit a total of 80 schools to account for attrition over the course of the study. Notably, we attend to three types of attrition in our study design. By recruiting 80 schools we allow for a rate of attrition among schools of around 10%.\(^3\) Additionally, our power analyses revealed that the power did not change appreciably when the number of teachers per school was cut in half. Finally, the student growth models addressing RQ2 are robust to missing data at level 1 of the analysis and thus growth models will be able to employ data from students with incomplete or complete data, accounting for the fact that we know there will be student attrition over the course of the study.

\(^3\) If necessary we can recruit additional schools from the district, to account for greater school-level attrition.
2.8 Cost-Benefit Analysis

We will also conduct a cost benefit analysis for the intervention using effect-size gains in student learning as the benefit. To gauge the cost of the IFL LW leadership training we will calculate the cost of a) training provided by IFL; b) time spent by four leadership team members outside the school in training sessions; and c) the cost of hiring substitute teachers for leadership team members with classroom teaching responsibilities. We propose to include benefits such as reduced grade failure, reduction in special education assignments, and projected learning gains as forecast by higher student test scores, in our analysis. As we know of no solid empirical data on the effects of a leadership development program on student achievement, we propose to compare the cost benefit of our intervention to the cost-benefit of class size reduction. Our reasoning is that there is relatively solid empirical evidence on the returns of reducing class size in terms of student achievement and we can calculate the cost of reducing class size in the district’s schools (Krueger, & Whitmore, 2001; Nye, Hedges, & Konstantopoulos, 2000).

2.9 Study Timeline and Products

The study timeline described below is tightly coordinated with The Learning Walk® training provided to treatment schools by the IFL (See Appendix B, Exhibit 5). We will pick our sample of treatment and comparison schools and finalize the School Staff Questionnaire (SSQ) survey in August 2010. We will develop our interview and observation protocols in September 2010 and select our purposeful sample of eight schools in early January 2011 based in part on an analysis of the SSQ baseline data.

The SSQ will be administered to both treatment and comparison schools through six evenly spaced administrations beginning in late September 2010 to track the measures of school practice longitudinally over the course of the study. The September 2010 administration will provide baseline measures of school practice, classroom practice, and leadership team members’ ability to critique classroom instruction and students’ learning based on their responses to video simulations. Measures of leadership team members’ ability to critique instruction and students’ learning based on the video simulations will be given in three administrations of the SSQ – September 2010 (baseline), September 2011, and April 2013.

Starting with the first training session in October 2010, IFL and district staff will keep detailed records of who attends each training session from each school. Beginning with the first LW school observations –in February 2011, we will collect artifacts generated as LW visits are performed in all treatment schools. We will continue to collect these artifacts through the end of the study. We will get the districts’ student achievement test scores. Baseline measures will be provided from the spring 2009 student scores and subsequent tests will be administered annually in the spring, providing a total of six years of student achievement data through Spring 2014. We will begin data collection for our embedded case studies in eight schools in January 2011 conducting interviews and observations in schools sampled from both the treatment and comparison schools. Data analysis involving quantitative, qualitative, and mixed method analyses will be ongoing over the three years of the study and extend into the fourth year after data collection is completed.

We will produce reports of our findings for publication in peer-reviewed journals such as *Journal of Research on Educational Effectiveness, Research Notes on*
Educational Effectiveness, Journal of Policy Analysis and Management, Educational Evaluation and Policy Analysis, American Educational Research Journal, Teachers College Record, and Educational Administration Quarterly. These reports each will focus on one or more of our eight hypotheses as well as reporting on other contributions of our study (e.g., using video simulations to measure leadership team members’ expertise in judging the quality of and critiquing classroom instruction). We will adapt at least two of these products into short, journalistic-style articles for placement in the periodicals of education associations such as the National Association for Elementary School Principals (NAESP) so as to ensure that our work reaches practitioners and policymakers. Project researchers will also present their findings at the annual meetings of national associations including American Educational Research Association (AERA), Association for Public Policy Analysis and Management (APPAM), National Association of Elementary School Principals (NAESP), Society for Research on Educational Effectiveness (SREE), and University Council for Education Administration (UCEA).

2.10 Personnel, Resources, & Advisory Committee

2.10.1 Personnel

James Spillane, Principal Investigator, is the Spencer T. & Ann W. Olin Professor in Learning and Organizational Change at Northwestern University where he teaches in the Human Development and Social Policy and Learning Sciences graduate programs. He is also a Faculty Fellow at the Institute for Policy Research and has an appointment in the Management and Organization Department at the Kellogg School of Management (Northwestern University). He directs Northwestern’s pre-doctoral Multi-disciplinary Program in Education Sciences funded by the Institute for Education Sciences. Spillane is also a co-PI on the Department of Education grant R305E040085, “Assessing the impact of principal’s Professional Development: An Evaluation of the National Institute for School Leadership.” He is author of Standards Deviation (Harvard University Press), Distributed Leadership (Jossey-Bass) and co-editor of Distributed Leadership in Practice (Teachers College Press). For the past decade, Spillane has engaged in an intensive study of school leadership focusing on the practice of school leadership in urban elementary schools and on the efficacy of a principal professional development program.

Brian Junker, Co-Principal Investigator, is a Professor of Statistics at Carnegie Mellon University and will direct the statistical analyses for this proposed research. His research interests include the statistical foundations of latent variable models for measurement, as well as applications of latent variable modeling in the design and analysis of standardized tests, smallscale experiments in psychology and psychiatry, and large-scale educational surveys such as the National Assessment of Educational Progress.

Richard Correnti, Co-Principal Investigator, is an Assistant Professor in the School of Education and Research Scientist at the Learning Research and Development Center at the University of Pittsburgh. His recent work includes evaluations of 1) CSR program effects on instruction, 2) the influence of content specific professional development on teacher practice, and 3) CSR program effects on student achievement. His current research efforts focus on evaluations of program interventions (e.g., a natural experiment
evaluation of a high school Science and Technology school with a lottery for admission), and also on different methods for examining instruction at-scale. For the proposed research, Dr. Correnti will be collaborating with Northwestern and Carnegie-Mellon researchers about data collection, designing survey instruments, analyzing data, and writing reports.

2.10.2 Resources

The Northwestern University School of Education and Social Policy (SESP) will provide facilities to support the proposed effort. The building is fully networked, including wireless connectivity. Network infrastructure is switched 10/100, relies upon a multiple T-1 connection scheme for outside access and is maintained by Northwestern University's Information Technology Services. Annenberg Hall houses state-of-the-art, networked “smart classrooms” which contain facilities for the use of computers, video and audio resources in instruction.

The University of Pittsburgh’s LRDC is a programmatic research institute for the scientific study of human learning, education, and training. The Institute for Learning, considered the country's best established and farthest reaching partnership between a university and urban school districts, is housed at LRDC. Established in 1995 as a response to district requests for help in meeting high academic standards, IFL has worked with more than 40 school districts since its founding, and currently has a membership of 14 districts in nine states. LRDC and Carnegie Mellon’s HCII department share leadership of the NSF-funded Pittsburgh Science of Learning Center, which specializes in conducting in vivo scientific research in schools.

2.10.3 Advisory Committee

A group of six leading scholars in school organizations, school leadership, social capital, and experimental design will consult on our research project. Anthony Bryk (President, Carnegie Foundation for the Advancement of Teaching), Ken Frank (Michigan State University), Larry Hedges (Northwestern University), Kenneth Leithwood (University of Toronto), Judith Warren Little (University of California, Berkeley), and Brian Rowan (University of Michigan) have agreed to join this group (see Appendix A).
## Appendix A

### Table A1: Sample Measures to be Adapted for the SSQ

<table>
<thead>
<tr>
<th>Measures Items</th>
<th>Reliability</th>
<th>Source**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Building a Vision (4 point scale, no priority to among highest three priorities)</strong>&lt;br&gt;Framing/Communicating broad goals for school improvement&lt;br&gt;Examining progress toward its school improvement goals&lt;br&gt;Setting explicit timelines for instructional improvement&lt;br&gt;Clarifying expectations or standards for students’ academic performance</td>
<td>.76-.82*</td>
<td>SII</td>
</tr>
<tr>
<td><strong>Involvement in Staff Development (4 point scale, no priority to among three highest priorities)</strong>&lt;br&gt;Examining and discussing exemplars of students’ academic work&lt;br&gt;Developing the staff development program in the school&lt;br&gt;Personally providing staff development in the school&lt;br&gt;Working on plans to improve the teaching of specific units/objectives</td>
<td>.66-.70*</td>
<td>SII</td>
</tr>
<tr>
<td><strong>Leader Observing Instruction (5 point scale, never to more than 10 times)</strong>&lt;br&gt;Teacher watched an instructional leader model instruction&lt;br&gt;An instructional leader observed my teaching techniques&lt;br&gt;An instructional leader watched my use of curriculum materials&lt;br&gt;Instructional leader studied student work</td>
<td>.69-.72*</td>
<td>SII</td>
</tr>
<tr>
<td><strong>Leader Advising Teachers on Instruction (5 point scale, never to more than 2 days per week)</strong>&lt;br&gt;Demonstrate instructional practices and/or use of materials in a classroom&lt;br&gt;Observe teacher trying new instructional practices or using new materials&lt;br&gt;Share information or advice about classroom practices with a teacher&lt;br&gt;Examine/discuss what students were working on during a teacher’s lesson</td>
<td>.81-.84*</td>
<td>SII</td>
</tr>
<tr>
<td><strong>Teacher Discourse about Teaching and Learning (5 point likert scale)</strong>&lt;br&gt;Share ideas on teaching with other teachers&lt;br&gt;Discuss with other teachers what you/they learned at a workshop or conference&lt;br&gt;Share and discuss student work with other teachers&lt;br&gt;Discuss particular lessons that were not very successful&lt;br&gt;Discuss beliefs about teaching and learning</td>
<td>.85</td>
<td>BASRC</td>
</tr>
<tr>
<td><strong>Teachers Observing Each Others’ Instruction (5 point likert scale)</strong>&lt;br&gt;Teacher observed another teacher teaching&lt;br&gt;Another teacher observed me teaching&lt;br&gt;Teacher taught with a colleague</td>
<td>.80</td>
<td>BASRC</td>
</tr>
<tr>
<td><strong>Instructional Leadership (4 point scale, strongly disagree to strongly agree)</strong>&lt;br&gt;Principal communicates a clear vision for our school&lt;br&gt;Instructional goals are made clear to staff&lt;br&gt;Principal presses teachers to implement what they have learned in staff development</td>
<td>.92</td>
<td>CCSR</td>
</tr>
<tr>
<td>Principal sets high standards for teaching</td>
<td>Principal sets high standards for student learning</td>
<td>Principal carefully tracks student academic progress</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Collective Responsibility: Peer-assessed (5 point scale, none to nearly all)</td>
<td>How many teachers in this school…</td>
<td>Feel responsible when students in school fail</td>
</tr>
<tr>
<td>Collective Responsibility: Self-assessed (4 point scale, strongly disagree to strongly agree)</td>
<td>There is really very little I can do to insure that most of my students achieve at a high level</td>
<td>My success or failure in teaching is due primarily to factors beyond my control</td>
</tr>
<tr>
<td>Collaboration - Standards for Learning and Content (5 point scale, never to more than 10 times)</td>
<td>Clarifying standards for student learning</td>
<td>Developing thematic units integrating instruction across curricular areas</td>
</tr>
<tr>
<td>Collaboration – Know Students’ OTL (4 point scale, strongly disagree to strongly agree)</td>
<td>Teacher knows content covered and instructional methods used by other teachers in school</td>
<td>Teacher knows what new students learned previously</td>
</tr>
<tr>
<td>Teacher-Principal Trust (4 point scale, strongly disagree to strongly agree)</td>
<td>Teacher are respected by other teachers.</td>
<td>Teachers in this school trust each other</td>
</tr>
</tbody>
</table>
Principal has confidence in the expertise of the teachers
I trust the principal at his or her word
It is okay in this school to discuss feelings, worries, and frustrations with the principal
The principal takes a personal interest in the professional development of teachers
The principal looks out for the personal welfare of faculty members

* Study of Instructional Improvement surveys were administered four times. The reliabilities reported in this table are the range of reliabilities reported in Rowan and Miller (2007).

** Abreviations are as follows: SII=Study of Instructional Improvement (See Rowan and Miller, 2007); BASRC=Say Area School Research Collaborative (See McLaughlin and Talbert, 2006); CCSR=Consortium for Chicago School Research (See, Sebring, Allensworth, Bryk, Easton, Luppescu, 2006)
Figure A.1: Power to detect teacher-level SP effects of LWT intervention, using a two-level model with Teachers nested within Schools (see main model in Section 2.5.1).

Figure A.2: Power to detect student-level achievement effects of LWT intervention, using a three-level model with Students nested within Teachers nested within Schools (see main model in Section 2.5.3).
Appendix B
The Learning Walk® Routine [EXHIBIT B-1]
The Learning Walk® routine is a highly structured set of activities for the observation and interpretation of teaching and learning based on the IFL’s Principles of Learning and focused on the instructional core—how teachers teach, how students learn, and what gets taught to whom. (Elmore, 1996) In keeping with the IFL’s fundamental commitment to effort-based education, The Learning Walk® routine also focuses on how a school is organized so effort creates ability. The routine consists of the following eight components practiced in a continuous cycle of observation and professional learning:

FOCUSING. The Learning Walk® leader or team uses the Classroom Instruction and Learning observation tool to specify an instruction/learning focus for its classroom observations and plans the classrooms to be visited and The Learning Walk® participants. The focus is based on current professional learning by the classroom teachers to be visited. Often, this professional learning has been planned in response to observations from a previous The Learning Walk® visit.

CONSULTATION. Once the focus for the walk is set, the leader informs teachers who will be visited of the visitation date and focus and asks for their guidance on what to observe within the chosen focus.

ORIENTATION OF WALKERS. Immediately before the walk, walkers receive updated information about the focus of and lens for the walk, including relevant data and materials provided by the teachers to be visited. At this step, walkers plan questions they might ask students that they believe will yield information pertinent to the focus.

CLASSROOM VISIT. The Learning Walk® school visit consists of three to five classroom visits, typically for about 10 minutes each. Different walkers make different observations, individually or in pairs. These include talking with students, examining classroom artifacts on the walls or boards or in student notebooks or portfolios, listening to teacher-student interactions, and listening to student-student interactions.

HALL TALK. After each classroom visit, the walkers have a brief conversation in the hall. The purpose is to check the accuracy of observations and ensure that all walkers are adhering to the focus and the frame for that particular walk. In addition to piecing together the evidence, walkers help each other understand what they have observed.

DEBRIEF. After all visits are completed, in preparation for discussing The Learning Walk® observations with the teachers whose rooms were visited, walkers meet to consolidate their observations and questions, looking for patterns across classrooms.


---
CONVERSATION WITH TEACHERS. Walkers discuss with teachers their observations and questions. They discuss possible next steps in professional learning and may consider a focus for a subsequent The Learning Walk school® school visit.

TEACHERS’ PLANNING. Teachers who have been visited discuss plans for their next step in collaborative learning. The principals, a coach, and/or a lead teacher are included in these planning sessions.