EXPLORING HOW SCHOOL INTRA-ORGANIZATIONAL MECHANISMS MEDIATE THE EFFECTS OF EXTERNAL INTERVENTIONS ON IMPROVING TEACHING AND LEARNING

By

Min Sun

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Education

ABSTRACT

EXPLORING HOW SCHOOL INTRA-ORGANIZATIONAL MECHANISMS MEDIATE THE EFFECTS OF EXTERNAL INTERVENTIONS ON IMPROVING TEACHING AND LEARNING

By

Min Sun

This dissertation collects three independent but interrelated studies exploring how school intra-organizational mechanism may mediate the impact of external interventions on improving teaching and learning. This first study examines how high-quality professional development (PD) can promote the diffusion of effective teaching strategies among teachers through collaboration. Drawing on sociometric data from a larger experimental evaluation study, this study shows that teachers' participation in professional development is associated with providing more help to colleagues on instructional matters. Moreover, the influence of professional development on teachers' instructional practice in writing spreads through this network of helping interactions in ways that augment the direct effect of participating in professional development, there can be spillover effects of professional development through collegial interactions. Evidence presented in this study will potentially help policymakers develop high-quality PD programs and distribute PD participants within schools to promote all teachers' performance.

The second study investigates the role of formal and informal leaders in supporting the diffusion of external reforms within schools. In the context of implementing reading policies as part of the No Child Left Behind (NCLB) 2001 legislation, this study aims to examine a) how formal and informal leaders promote instructional changes through professional interactions with teachers; and b) which types of instructional practices are most responsive to which types of leaders. I analyze longitudinal data concerning both professional interactions about teaching

reading and instructional practices of teachers and leaders in nine K-8 schools in a single state. I find that formal leaders convey normative influence on general teaching practices such as setting standards, selecting materials, and assessing students, while informal leaders convey normative influence on specific pedagogical practices of teaching basic reading skills. Findings contribute to the theoretical and methodological development of both distributed leadership and policy implementation within schools. Moreover, this study suggests the importance and several strategies for developing a strong instructional leadership team that recognizes the complementary influences of formal and informal leaders.

The purpose of the third paper is to investigate the potential of using multilevel item response theory to estimate the depth of teacher interactions under the ego-centric framework, defined as the propensity of endorsing collaborative relations with regard to mathematics instruction. Using empirical data from a larger study of Middle School Mathematics and the Institutional Setting of Teaching, beyond estimating item rareness parameters (fixed effects) and the depth of interaction for each tie (random effects), this study also develops methods to gauge item goodness-of-fit and information function to examine the quality of social network survey instruments under a multilevel framework. Finally, this study demonstrates the possibility to incorporate predictors in the measurement model to investigate Differential Item Functioning (DIF) and explanatory research questions. The methodological development in this paper significantly contributed to the growing popularity of using network studies to inform educational policy and practices by providing a psychometrically sound measure of teacher interaction on professional matters.

To Drs. Kenneth A. Frank and Peter Youngs who shepherded my academic career from its infancy to my acceptance as an educational researcher

ACKNOWLEDGEMENTS

I am heartily thankful to my advisor in Measurement and Quantitative Methods program, Dr. Ken Frank, and to my advisor in Educational Policy program, Dr. Peter Youngs. Without their multiple years of consistent and extensive support, this dissertation would not be possibly written. Moreover, it is very easy for a doctoral student who is pursuing a dual-major degree to become a "generalist" but not a "specialist". I very appreciate the collaboration between Ken and Peter with regard to advising my coursework, research, and teaching activities. Without their collaboration, I would not have been able to develop a coherent set of knowledge and skills, as well as a research agenda centering on the issues of how school intra-organizational mechanisms may mediate the individual outcomes of educational reform and policy.

Dr. Ken Frank has been my advisor and mentor since 2007. He has given me enormous opportunities to learn and grow, has provided the necessary encouragement and reinforcements for me to complete all coursework and exams required by both programs, and has paved the way of my career development by helping me establish my professional networks. He brought me into these three large-scale studies, from which this dissertation draws data. I am indebted to him for all of his support.

Dr. Peter Youngs has been my advisor and mentor since 2006 when I just landed in the United States from Beijing, China. He discovered my dual interests in both educational policy and methodology and encouraged me to pursue these dual interests. Moreover, with his financial support and professional guidance, I was able to collect my own data and publish my first paper in English about how districts evaluate principals to promote learning-centered leadership activities. Besides, I am grateful to him for keeping me on track with the program agenda of Educational Policy.

v

I also owe my deepest gratitude to my dissertation committee, Drs. Bill Penuel, Gary Sykes and Mark Reckase. Dr. Penuel has taught me many things beyond what graduate schools can offer, such as the ways of engaging stakeholders in the evaluation research. His interdisciplinary expertise, coupled with his sharp sense of conducting empirical research, has greatly shaped my dissertation and other publications. Dr. Sykes' insight on educational policy helped me clarify the complicated issues with which I am untangling in the first two sub-studies of this dissertation. Dr. Reckase's expertise in educational measurement has extensively shaped the third sub-study. Dr. Susan Printy and Dr. Spyros Konstantopoulos, although not officially named as my advisors or dissertation committee members, have provided essential support and encouragement at some key moments, and I appreciate very much their time and their concern.

I would like to acknowledge the generous financial support from the American Educational Research Association (AERA) dissertation grant program in conjunction with National Science Foundation. Moreover, Dr. Jacquelynne Eccles, who serves as the representative of the AERA grant governing board members and has supervising this dissertation, has helped me make the results more accessible to policy makers and funders. I would also like to acknowledge the enormous support from Dr. Alix Gallagher at SRI International, Dr. Linda Friedrich at the National Writing Project, and colleagues at Vanderbilt University, such as Dr. Paul Cobb, Dr. Tom Smith, Annie Garrison, and Dr. Christy Larson. Without their support, the key analysis in this dissertation would not have been possible.

Lastly, I would like to thank the long-standing support of my family and friends. My husband, Jian (Kevin) He, has been ceaselessly supporting me in the last five years and has been my best friend since we met. I certainly cannot enumerate all of the support from my parents and my parents-in-law. Moreover, my dearest friend, Mary Mason has selflessly shared her rich

vi

professional and life experiences with me. Thanks go as well to fellow (and former) graduate students, including Dr. Yongmei Ni, Yisu Zhou, and Andrew Saultz.

PREFACE

This research is supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation under Grant #DRL-0941014. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies.

LIST OF TABLES	xii
LIST OF FIGURES	xiii
CHAPTER 1: THE SIGNIFICANCE AND NOVELTY OF THIS DISSERTATION	1
Practical Significance	1
Theoretical Significance	2
Methodological Significance	4
References	7
CHAPTER 2: SHAPING PROFESSIONAL DEVELOPMENT TO PROMOTE THE	
DIFFUSION OF INSTRUCTIONAL EXPERTISE AMONG TEACHERS	9
Introduction	9
Literature Review	10
Effective Features of Professional Development	10
How Spillover Effects of Professional Development Occur through Collegial Interaction	ıs?
	12
The Development of the Current Study	14
Sample	16
Measures	19
Dependent Variables	19
Focal Predictors	20
Analytic Strategies	22
Estimation of the Number of Colleagues Helped with Teaching Writing	23
Estimation of How PD Shapes Instructional Practices through Collegial Interactions	26
Quantifying Robustness of Inferences	29
Results	30
Descriptive Statistics of PD features in Both Partnership (Treatment) and Delayed	
Partnership (Control) Schools	30
Effects of PD Features on the Number of Colleagues Helped With Teaching Writing	31
Effects of PD Features on Instructional Practices through Professional Help	34
Discussion	38
Substantive Interpretation	38
Methodological Interpretation	41
Limitations	42
Policy Implications	43
Conclusion	44
References	56
Appendix 2.A	47
Appendix 2.B	48
CHAPTER 3: HOW EXTERNAL INSTITUTIONS PENETRATE SCHOOLS THROUGH	_
FORMAL AND INFORMAL LEADERSHIP?	56
Introduction	64

TABLE OF CONTENTS

The Distinctive Influences of Formal and Informal Leaders on Instruction	66
Formal Leaders' Influence	66
Informal Leaders' Influence	67
Hypotheses Regarding the Motivations of the Distinctive Influences in the Implementati	on of
Accountability Reform	68
Sample and Measures	71
Sample	71
Measures	72
Data Analysis	76
Results	79
Estimating Effects on General Practices of Implementing NCLB- Related Standards,	
Curricula and Assessments	79
Estimating Effects on Specific Pedagogical Practices of Teaching Basic Reading Skill	s 80
Discussion	81
Theoretical Implications	82
Practical Implications	84
Limitations	85
Conclusion	86
References	95
Appendix 3	88
CHAPTER 4: THE USE OF MULTILEVEL ITEM RESPONSE THEORY MODELING ESTIMATE PROFESSIONAL INTERACTIONS AMONG TEACHERS	ГО 94 101
Introduction	101
Multilaval Itam Dasponsa Theory Models	102
Instrument and Semple	105
Models	107
Maggurement Model	108
Prediction Model	100
Paculte	111
Testing Pasch Model Assumptions	113
Item Information or Goodness of Fit Indices under Multilevel Framework	113
Item Pareness	117
The Distribution of Propensity of the Interactions	110
Instrument Diagnosis: Calculating Information Function	110
Differential Item Function (DIF)	117
Prediction Model Estimates	117
Discussion	120
Conclusion	121
References	146
Appendix 4 A · Instrument	140
Appendix 4 B. Figures	120
Appendix 4 C. Tables	135
	155
CHAPTER 5: SUMMARY AND CONCLUSIONS OF THIS DISSERTATION	146

References 154

LIST OF TABLES

Table 2.1 School Characteristics in 2008-09.	. 18
Table 2.2 Teacher Characteristics in 2008-09 School Year	. 18
Table 2.3 Estimates of the Contribution of PD Features To Year-2 Instructional Practices	. 26
Table 2.4 Pearson Correlation between PD Features	29
Table 2.5 Descriptive Statistics of PD features in both Partnership (Treatment) and DelayedPartnership (Control) Schools in 2009-10	31
Table 2.6 Estimated Effect of PD Features on the Number of Colleagues Helped With Teaching Writing	ng 31
Table 2.7 Estimated Effects of PD Duration on Instructional Practices	. 35
Table 2.8 Estimated Effects of PD Content on Instructional Practices	35
Table 2.9 Estimated Effects of PD Format on Instructional Practices	36
Table 3.1 School Demographic Information in 2007-08.	. 72
Table 3.2 Teacher Demographics from 2008 Survey.	72
Table 3.3 Demographic Characteristics of Formal and Informal Leaders	70
Table 3.4 Estimating General Practices of Implementing NCLB-Related Standards, Curricula, and Assessments in 2008.	80
Table 3.5 Estimating Specific Pedagogical Practices of Teaching Basic Reading Skills in 2008.	.81
Table 4.1 Item Descriptions 1	108
Table 4.2 Ego's Characteristics in the 2008-09 School Year	108
Table 4.3 Tetrachoric Correlation Matrix 1	114
Table 4.4 Error Variances and R-square by Items 1	115
Table 4.5 Compare Goodness of Fit between Two-Factor Model and One-Factor Model Using CFA	g 115
Table 4.6 Factor Loadings	115

Table 4.7 One- and Two-Parameter Models	117
Table 4.8 Item Goodness-Fit Indices under Multilevel Framework	118
Table 4.9 Item Rareness Estimates	118
Table 4.10 Differential Item Function (DIF) Parameter Estimates	120
Table 4.11 Fixed Effects of Prediction Model	120

LIST OF FIGURES

Figure 2.1 The Theoretical Framework of How Features of PD Affect the Dynamic of Diffus Expertise among Teachers	sing 15
Figure 4.1 Egocentric Network Structure	102
Figure 4.2 Egocentric Network Data Structure	103
Figure 4.3 The Ten Largest Eigenvalues in Order of Size	115
Figure 4.4 Graphical Comparison of Two-and One- Parameter model	117
Figure 4.5 The Distribution of Propensity of the Interactions	119
Figure 4.6 The Distribution of Information against the Propensity of the Interactions1	119

CHAPTER 1: THE SIGNIFICANCE AND NOVELTY OF THIS DISSERTATION

This dissertation reports on three independent but interrelated studies exploring how school intra-organizational mechanisms may mediate the impact of external interventions on improving teaching and learning. The first sub-study (Chapter 2) investigates the ways of shaping professional development programs to promote the diffusion of instructional expertise through teacher interactions; the second sub-study (Chapter 3) investigates how external institution may penetrate schools through formal (such as principals, coaches, or department chairs) and informal (such as regular teachers who do not have any leadership roles) leadership; and the third study (Chapter 4) applies multilevel item response theory to estimate the depth of professional interactions among teachers. This chapter aims to explain how these three substudies collectively contribute to the development of current educational policy and practice, the theoretical understanding of teacher learning and institutional diffusion, and the methodological enhancement in measuring teacher interactions and conducting rigorous data analysis.

Practical Significance

From the No Child Left Behind Act of 2001 (NCLB) to the Obama administration's blueprint for reauthorization of the Elementary and Secondary Education (ESEA), policy efforts for improving all students' learning have been intensified in recent legislation. Achieving such an ambitious goal depends on effective policy implementation, and a key aspect of implementation is improving teaching (Cohen, Raudenbush, & Ball, 2003). Not surprisingly, then, there are many debates about how to support teachers' professional learning and instructional change, of which three of the primarily supports are professional development, instructional leadership, and peer collaboration. To inform the current practice and policy of supporting improvement in practice, the first two sub-studies of this dissertation extend beyond

the evaluation of the overall impact of a practice and focus on examining how intraorganizational mechanisms via professional networks can help explain and predict the outcomes. The development of explanatory and predictive theories of educational processes and mechanisms would greatly contribute to the design, implementation, and evaluation of educational interventions. Without an evidence-based theory of educational processes and mechanisms, we would not be able to probe the inside of the "black box" of school practice and understand whether and how the evidence-based effectiveness can be generalizable to new settings or different populations (Society for Research on Educational Effectiveness, 2011).

Theoretical Significance

Researchers have posited two broad views of how teachers improve their classroom practices. One body of scholarship views teachers as individual learners and emphasizes the knowledge and information available to teachers and the role of professional development in supporting teachers' accrual of this knowledge and information (e.g., Ball & Cohen, 1996; Grant, Peterson, & Shojgreen-Downer, 1996; Senger, 1999). The other body of scholarship focuses on the structural or organizational influences on school norm and classroom instructional practices (e.g., Rowan & Miller, 2007). This dissertation engages both of these bodies of scholarship. Specifically, the first sub-study of this dissertation stems from the first body of scholarship and investigates the influence of high-quality features of professional development programs on teachers' instructional practice in writing, while the second sub-study of this dissertation contributes to the second body of scholarship and investigates the ways in which formal and informal leadership influences teachers' instructional practices of teaching reading.

This dissertation extends beyond viewing the improvement of teaching from either an individual or organizational perspective. For example, in Chapter 2 of exploring the effects of

professional development, besides estimating the direct impact of high-quality features of professional development, I examine the extent to which the influence of professional development on teachers' instructional practice in writing spreads through the network of helping interactions in ways that augment the direct effect of participating in professional development. Beyond viewing teaching as an isolated practice, this study acknowledges the interactions among teachers and views teaching as a set of collective and shared activities (Cobb, 2003; Coburn, 2001). Therefore, the impact of professional development programs for individual teachers can expand beyond individuals: it can spill over to other teachers through these professional interactions and shared activities.

Moreover, the second sub-study extends the second body of scholarship. This study finds that formal leaders convey normative influence on general teaching practices such as setting standards, selecting materials, and assessing students, while informal leaders convey normative influences on specific pedagogical practices of teaching basic reading skills. These findings elaborate on Spillane et al (2006)'s concept of distributed leadership by identifying these separate yet complementary influences on the change of instructional practices from formal and informal leadership. Moreover, these findings contribute to our understanding of leadership functions in the implementation of external reform by probing into the mechanism of how influence on instructional practices had occurred. The common contribution of these first two sub-studies is the theoretical development of institutional diffusion. External reform or policies, which constitute the institutional setting in which teachers work, do not uniformly diffuse within schools (Frank, Penuel, Sun, Kim, & Singleton, under review). External interventions often are adopted by some individuals, and then channeled through the grids of school formal and informal structure to enact influences on all school staff. These intra-organizational mechanisms

through leadership and collegial interactions can mediate the implementation and impact of these external interventions on instructional practices.

Methodological Significance

This dissertation has two main methodological contributions to the study of teacher interactions and the implementation of educational policies. The first is the development of a measure of professional interactions among teachers that has desirable psychometric properties. The collection of social network data allows us to ask many important questions for educational leadership and policy, such as "how do interventions affect individual outcomes?", or "how and why do formal and informal interactions among school staff medicate the successful implementation of external interventions?", or "who are the key actors in the mediation process?", or "how can external reforms intervene this mediation process?" These questions are intended not only to explain these intra-organizational mechanisms but also to predict the impact of external interventions on such mechanisms. Collecting social network data can help researchers to develop direct measures to investigate these questions and has been used in several large educational studies, such as the National Evaluation of Writing Project Professional Development (Gallagher et al., 2010), the analysis of the "implementation gap" in high school reforms (Supovitz et al., 2008), and the examination of the institutional settings for mathematics teaching in urban middle schools (Cobb & Smith, 2008). The growing popularity of using social network data demands the development of a psychometrically sound measure of teacher interaction. In the third study elaborated in Chapter 4, I apply multilevel item response theory models to social network data and estimate the propensity of endorsing collaborative relations with regard to teaching mathematics, on a continuum. These models also allow one to gauge the

goodness-of- fit of each item and the information function of the whole instrument to much better understand the properties of the instrument.

Second, the first two sub-studies that address important policy and leadership questions aim to establish confidence for drawing causal inference. These two studies draw on data from two large-scale longitudinal studies: *SRI International's National Evaluation of Writing Project Professional Development* by implementing random experiments, and the study of *Analyzing the Flow of Network-Embedded Expertise in Schools: A Longitudinal Study of Individual and Organizational Change*. The high quality data allow me to conduct rigorous longitudinal data analysis to reduce the impact of selection bias and confounding variables by 1) including prior measures of dependent variables in the modeling; 2) using school fixed effects to control for disparities across schools; 3) quantifying the robustness of the inferences about the estimates in terms of the extent to which bias due to any omitted confounding variables would have been necessary to have made these inferences invalid.

In summary, this dissertation is dedicated to providing more rigorous empirical evidence to advance our understanding of how the effectiveness of instructional reforms is mediated by the formal and informal settings in which teachers' instructional practices are situated. This dissertation will inform the investment in professional development programs, formal and informal leadership, and teacher professional communities to support instructional improvement, and will explore how to leverage this interwoven complexity in school settings to provide coherent guidance on instructional reforms.

REFERENCES

REFERENCES

- *Building an Education Science: Investigating Mechanisms.* (2011). Retrieved from Society for Research on Educational Effectiveness, http://www.sree.org/conferences/2011/
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is--or might be--the role of curriculumn materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6-14.
- Cobb, P. & Smith, T. (2008). The challenge of scale: Designing schools and districts as learning organizations for instructional improvement in mathematics. In K. Krainer, & T. Wood (Eds.), *International handbook of mathematics teacher education: Vol. 3. Participants in mathematics teacher education: Individuals, teams, communities and networks* (pp. 231-254). Rotterdam, The Netherlands: Sense.
- Cobb, P., McClain, K., Lamberg, T. d. S., & Dean, C. (2003). Situating teaches' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*(6), 13-24.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, Instruction, and Research. *Educational Evaluation and Policy Analysis*, 25(2), 119-142.
- Gallagher, H. A., Woodworth, K. R., Bosetti, K. R., Cassidy, L., McCaffrey, T., Yee, K., et al. (2010). National Evaluation of Writing Project Professionall Development Year 4 Report: SRI International.
- Grant, S. G., Peterson, P. L., & Shojgreen-Downer, A. (1996). Learning to Teach Mathematics in the Context of Systemic Reform. *American Educational Research Journal*, 33 (2), 502-541.
- Grant, S. G., Peterson, P. L., and Shojgreen-Downer, A. "Learning to Teach Mathematics in the Context of Systemic Reform." *American Educational Research Journal*, 1996, 33(2): 502–541.
- Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal*, 44(2), 252-297.
- Senger, E. (1999). Reflective reform in mathematics: The recursive nature of teacher change. Educational Studies in Mathematics, 37, 199-221.
- Spillane, J. P. (2006). Distributed leadership. San Francisco, CA: Jossey-Bass.

Supovitz, J., & Weinbaum, E. H. (2008). Reform Implementation Revisited. In J. Supovitz & E.
H. Weinbaum (Eds.), *The implementation gap: Understanding reform in high schools*.
New York: Teachers College Press.

CHAPTER 2: SHAPING PROFESSIONAL DEVELOPMENT TO PROMOTE THE DIFFUSION OF INSTRUCTIONAL EXPERTISE AMONG TEACHERS

Introduction

Effective teaching matters for student learning. From the No Child Left Behind Act of 2001 (NCLB) to the Obama administration's blueprint for reauthorization of the Elementary and Secondary Education (ESEA), policy efforts for improving all students' learning have been intensified in recent legislation. Achieving such ambitious goals depends on effective policy implementation, and a key aspect of implementation is improving teaching (Cohen, Raudenbush, & Ball, 2003). Professional development (PD) has been assumed to be a promising avenue for improving teacher quality (Correnti, 2007), particularly in the long run (Hoxby, 2003a). To inform the theory about how PD can promote instructional improvement, this study extends beyond the direct impact of a PD program and focuses on examining how the school intraorganizational mechanisms via professional networks can augment the PD outcomes. The development of explanatory and predictive evidence of these mechanisms would greatly contribute to the design, implementation, and evaluation of PD programs (Society For Research on Educational Effectiveness, 2011). With an evidence-based theory of educational processes and mechanisms, we can probe the inside of the "black box" of school practice and understand better what constitutes effective PD programs and how teachers can learn best in the local settings where they are situated (e.g. Cobb, McClain, Lamberg, & Dean, 2003).

Previous large-scale evaluation studies have focused on changes in participants' instructional practices as direct measures of the outcomes of PD programs (reviewed in the next subsection). None of them have examined the *spillover effect* of PD, as we define it here, which refers to effects of PD availability in a school that shapes instructional practice above and

beyond direct participation in PD. Where some studies focus on how curriculum materials can enhance the impact of PD on practice (e.g., Penuel & Gallagher, 2009), few studies have explored how collegial interactions can enhance the impact of professional development. In this study, we assess PD spillover effects in two measures. The first one is the increase in the number of colleagues a teacher helped after participating in high-quality PD (Frank et al., 2008); that is, I examine whether PD makes participants more likely to become the "go to" experts for professional matters. The other measure is the extent to which colleagues' instructional practices have been improved after receiving help from PD participants (Frank et al., 2004; Frank, Zhao, Penuel, Ellefson, & Porter, 2011).

Drawing on longitudinal data collected from a larger experimental evaluation of the National Writing Project's partnership activities, a study called the National Evaluation of Writing Project Professional Development (WPPD), this paper aims to analyze the spillover effects of PD and how high-quality PD features shape knowledge diffusion among teachers. To inform the theoretical development of this study and data analysis, I first review literature examining the direct effects of PD features on teachers' classroom practices. I then elaborate how the spillover effect could occur through collegial interactions and the significance of examining PD participants' spillover effects.

Literature Review

Effective Features of Professional Development

Studies have examined various PD programs, with some consensus on what constitutes high-quality PD. First, as opposed to a one-time presentation or one-day workshop, PD should be sustained over time (e.g. Correnti, 2007; Darling-Hammond et al., 2009; Yoon et al., 2007; Garet et al., 2001; Newmann et al., 2000). However, there is no exact number of sufficient PD

hours. For example, the average number of contact hours was 25 in one year in the Eisenhowerassisted PD (Garet et al., 2001). In comparison, Yoon et al. (2007) synthesized nine welldesigned studies and concluded that intensive professional development offered for an average of 49 hours in a year boosted student achievement by approximately 21 percentile points.

Second, the content of PD should represent effective instructional practices, be aligned with curricula and assessments in schools, and focus on specific subjects that teachers are teaching (Cohen, Raudenbush, & Ball, 2002; Corcoran, 1995; Correnti, 2007; Garet et al., 2001). Empirical studies have shown that such content-driven PD has significant and positive associations with teachers' self-reported increases in knowledge and skills and changes in classroom practices (Cohen & Hill, 2001; Desimone et al., 2002; Garet et al., 2001; Penuel et al., 2007). I also anticipate that the degree to which PD develops teachers' collaborative learning skills can positively predict the impact of PD on both teaching and learning (Cordingley, Bell, Rundell, & Evans, 2005).

Third, the delivery method or formats of activities used to engage teachers in the learning process also matters. Research results suggest that job-embedded PD is effective, such as when a PD provider observes a teacher during a normal class instruction or when teachers discuss with coaches their lesson plans (e.g., Darling-Hammond et al., 2009). Moreover, PD activities that involve active learning, such as small group discussion, show more effects on instructional practices than lectures. Collective participation of teachers from the same school, grade, or subject is also an evidence-based practice (Darling-Hammond et al., 2009; Desimone et al., 2002).

Studies have connected these PD features to two types of direct measures of outcomes: PD participants' content knowledge and their instructional practices. However, to date no studies

have probed into the effect of the provision of PD to some teachers on the performance of other teachers in the school who may or may not have direct participation. Yet we know that characteristics of schools which can moderate the efficacy of PD, such as enrollment size, the socioeconomic status of the student population, and urbanicity, etc., may affect teachers' actions as well (Darling-Hammond & McLaughlin, 1995; Firestone, 1996; Newmann et al., 2000; Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2009; Stein & Lane, 1996). Besides these contextual moderators, as I describe below, there is evidence to suggest that intra-organizational processes could augment the direct effects of PD and thus merit more systematic study. *How Spillover Effects of Professional Development Occur through Collegial Interactions?*

A number of studies show that teachers' instructional practices are shaped by interactions with colleagues and participation in collaborative learning communities (e.g. Bryk & Schneider, 2002; Jackson & Bruegmann, 2009; McLaughlin, 2006; Newmann et al., 2000). Teachers often help one another with instructional matters, and often the provision of help involves knowledge transfer among teachers. Peers' knowledge and instructional expertise can be a major source of professional growth for teachers when interactions involve activities that give rise to learning opportunities. With regard to instructional changes as outcomes, teachers benefit from exposure to information that is embedded in classroom practices and peers possess this kind of instructional expertise (e.g., Darling-Hammond & McLaughlin, 1995; Webster-Wright, 2009). This expertise diffuses when teachers interact and collaborate with each other to address commonly-identified classroom problems (Penuel et al., 2006). Grounded on the core principle that informed and effective teachers can be successful teachers and partners of their colleagues, many reform programs, including the National Writing Project and the Coalition of Essential Schools, have focused on promoting teacher collaboration and professional learning

communities as a way to improve teacher quality and school capacity (Lieberman & Wood, 2002; Rowan & Miller, 2007).

Such reform programs often encourage teachers to provide help to others and thus to function as teacher leaders. These teacher leaders contribute to the successful implementation of reforms by working with other teachers on collectively interpreting policy messages and developing specific guidance of how to teach in classrooms (Coburn, 2001). They may also lead other teachers to lobby for shared resources, increasing the amount available to each teacher (Jackson & Bruegmann, 2009). Furthermore, our prior study provided evidence that in the implementation of external interventions, the normative influence of these teacher leaders on the core of classroom teaching may surpass the impact of formal leaders such as principals, department chairs, and coaches (c.f. Crowther, Ferguson, & Hann, 2009; Supovitz, 2008; Supovitz, Sirinides, & May, 2010), because compared to formal leaders, teacher leaders who are engaged in classroom themselves have specific pedagogical knowledge of what to teach and how to teach (Sun et al., 2010).

Moreover, collegial interactions involve collaboration, which has been acknowledged as a crucial school condition for teachers to retain enthusiasm for teaching. Large-scale surveys have shown that in-service teachers place a high priority on collaboration with school leaders and other teachers when they make a decision of whether to continue to teach or stay in the current school (e.g. Ingersoll & Smith, 2004; MetLife American Teacher Survey, 2009; North Carolina Teacher Survey, 2008). The provision of help and transfer of knowledge contribute to the store of "social capital" on which schools can draw to attract, retain, and develop high-quality teachers (Coleman, 1988; Frank et al., 2008).

The Development of the Current Study

Research provides evidence concerning the direct effects of PD features on changing participants' knowledge and instructional practices. However, almost none of the prior studies have explored how PD's effectiveness can be enhanced by shaping knowledge diffusion in the school community and by changing relational dynamics in ways that augment direct effects of PD. After Jackson and Bruegmann (2009) who present evidence of effects of teachers' peers on teachers' student achievement, we call this a *spillover* effect of PD. The spillover mechanism through professional interactions and the spillover effects of PD participants on the performance of others in the school is not currently well explored. In the effort to close the literature gap, I specifically ask two questions with regard to the measures of spillover effects of PD that I elaborated at the beginning:

1) How do PD features, such as the duration (contact hours), content (substance), and format (delivery methods), affect the number of colleagues a teacher helps with teaching writing?

2) How do effective PD features promote desired instructional practices through shaping knowledge diffusion within schools?

I articulate hypotheses associated with each research question by drawing on the theory of innovation diffusion. In brief, literature on innovation diffusion examines "the process in which the message of new ideas is communicated through certain channels over time among members of a social system" (Rogers, 1995, p.5). At a given point, new information may be channeled through social networks to only certain potential adaptors (Abrahamson & Rosenkopf, 1997). If members of the organization can access this information by interacting with these particular people, they may be influenced by them, adopt the innovation and thus change their

social behaviors according to the new expectations (Frank & Fahrbach, 1999; Monge, Cozzens, & Contractor, 1992). This internal diffusion process can be affected by modifying the information sources and communication channels (Nilakanta & Scamell 1990). Applying this theory to investigate the central research questions of this study, PD as an *external pressure or a knowledge source affects teachers' instructional practices through shaping the internal knowledge diffusion within school organizations (as illustrated by Figure 2.1).* In what follows, I untangle the dynamic of how instructional expertise potentially spreads among teachers and how PD may affect this mechanism.

Insert Figure 2.1 in Appendix 2.A about Here

Hypothesis 1: Teachers would more likely to provide help with writing instruction if they had participated in high-quality PD programs. The quality of PD programs can be measured on three features: the duration of participation (contact hours), content (substance), and format (delivery methods).

The desire for certain information leads teachers to seek help from those who possess this information (Knott, 2003). PD provides teachers with new sources of information, such as content knowledge, knowledge about student learning, and skills for communicating with students and colleagues, as well as state requirements on curriculum and assessment. Through involvement in various PD programs, teachers absorb, adopt and implement new sources of information offered by PD into the expertise of teachers themselves. The knowledge and resources are crucial for teachers to succeed in their job and thus may be desired by others who want to show

effectiveness (Youngs et al., in press). PD participants may also become help providers because PD highlights their roles as content experts. Effective PD can shape or restructure knowledge diffusion within a given school by determining whom teachers interact with and how many resources they can seek from others. Moreover, if teachers have been involved in PD programs that feature active learning, collective participation, and leadership, they will be better prepared to engage in deep collaboration than peers who have not been involved in such PD programs (Lieberman & Wood, 2007). Finally, the long duration of PD participation increases the amount of opportunities for participants to learn and interact with PD providers and with each other.

Hypothesis 2: The expertise that teachers gain from participation in PD will spread out to colleagues through the provision of help and thus change colleagues' instructional practices.

Exposure to new information or resources by interacting with colleagues may change teachers' own behaviors as a result of influence by colleagues with whom they interact (e.g. Abelson & Bernstein, 1976; Burt, 1982). The extent to which a teacher is influenced by interacting with others is a function of the content and frequency of their interactions, as well as the available expertise of their colleagues (Frank, Zhao, & Borman, 2004). Content knowledge and communication skills gained from PD could thus help teachers to disseminate their influence. Thus, PD not only affects participants' instructional practices, but also others with whom participants offered or have offered professional help. Such internal dynamics of knowledge diffusion shape the impact of PD on the school's average performance across all classroom teachers.

Sample

This study draws on data from a larger study of the evaluating the impact of the National Writing Project on teachers' instructional practices. Researchers at SRI International randomly

assigned 39 schools that served middle grades students to one of the two experimental conditions: 20 schools were assigned to the partnership (treatment) condition and 19 schools were assigned to the delayed partnership (control) condition. To increase the chances that the groups were comparable on key contextual factors, researchers recruited schools in pairs within Local Writing Project sites¹, restricting each site to a maximum of two pairs, and assigning one school in each pair to each condition.

The experiment started in 2007-08 with a baseline year in which no schools participated in any new writing professional development. In the partnership condition, schools agreed to spend the three years after the baseline year planning and implementing a partnership with a Local Writing Project site. These partnership strategies were co-designed and included a range of school-based in-service offerings, in-school and in-classroom coaching, and structured and unstructured opportunities for teachers to work with peers in professional learning communities. In the delayed partnership condition, except for district and state required programs, schools were asked to refrain from participating in any additional schoolwide professional development related to writing for the baseline year (2007-08) and three subsequent years (2008-09 through 2010-11).

These 39 schools were located in 14 Local Writing Project sites across the nation. Schools in the two experimental conditions were comparable with regard to various background characteristics (Gallagher et al., 2009). As demonstrated in Table 2.1, in the partnership schools, the average enrollment size was 669 with a standard deviation of 368, compared to the average enrollment size of 564 with a standard deviation of 269 in delayed partnership schools. The average percentage of students who were eligible for free- and reduced-price lunch was about

¹ Local Writing Project sites are university-based institutes that provide each school in the treatment condition with customized writing professional development programs.

44% in partnership schools and about 53% in delayed partnership schools. The majority of students were White in both experimental conditions. The average pupil-teacher ratio was around 15 to 1. The schools had an average of 45 full-time equivalent (FTE) teachers, about four of whom taught English language arts (ELA).

Insert Table 2.1 in Appendix 2.B about Here

The differences in teacher characteristics between partnership and delayed partnership schools were not statistically significant, as shown in Table 2.2. In the first year of implementation in 2008-09 and across all schools, teachers, on average, had 13 years of teaching experience with a standard deviation of 9.7. On average, they had taught in current schools for eight years. More than 90% of the teachers had a Bachelor's or Master's degree in both experimental conditions. About 5% of teachers had an education specialist degree or a professional diploma based on at least one year's work past the Master's degree. Few teachers had doctorates.

Insert Table 2.2 in Appendix 2.B about Here

The larger study invited all credentialed staff (except for principals) in the 39 schools to take annual surveys, which included questions about PD experience, teachers' professional networks, instructional practices, school contexts, and individual background information. The measures in this study were derived from the annual teacher surveys collected in spring 2008 (Year 1, Baseline), spring 2009 (Year 2, the first year of implementation), and spring 2010 (Year

3, the second year of implementation). In what follows, I briefly summarize how the measures were constructed and which wave(s) of data were used.

Measures

Dependent Variables

The Number of Colleagues Helped with Teaching Writing in Year 3 (2009-10 School Year)

In the 2010 spring survey, teachers were asked to nominate other teachers who had been helpful with teaching writing. The dependent variable is then simply the total number of other teachers who nominated a teacher as helpful. Thus, if Lisa were nominated as having provided help to Joe, Sue, and Bob, then Lisa's value would be 3, because she was nominated by three other teachers. In this measure, I followed Frank's (2004; 2008) work of emphasizing the import of obtaining the measure from the recipients of help rather than help providers. That is because expertise with regard to instructional matters likely has been transferred only if the recipient indicates such, regardless of reports of those who originally possess expertise and attempt to transmit knowledge (Hansen, 1999).

Instructional Practices in Year 3 (2009-10 School Year)

The survey asked teachers to report on frequency with which they engaged in researchbased instructional practices in writing. The items for these practices were drawn from metaanalyses conducted by Graham and Perin (2007a; 2007b; 2007c) that focused on teaching strategies targeting middle and high school students. Their meta-analysis included only experimental and quasi-experimental intervention studies. The strategies identified as effective that were included in survey items were: strategy instruction focused on planning and revising writing, summarization instruction, collaborative writing with peers, establishing specific goals for writing, providing good writing that served as models for students, engaging students in prewriting exercises to gather and organize ideas, and developing students' abilities to selfmonitor their writing. I aggregated two measures of high-quality writing instruction that drew from specific survey items:

The breadth of writing purposes taught in Year 3: In spring 2010 survey, each teacher was asked to rate how often they had students engage in writing for purposes "To reflect on an experience or topic (e.g., journaling)", "To express themselves creatively (e.g., a poem, story, or play)", " To recount a story or event through narrative", "To describe a thing, place, process, or procedure (e.g., an essay, lab report, or descriptive response)", "To explain a concept, process, or relationship (e.g., comparison/contrast, problem/solution)", "To make an argument intended to persuade others", "To gain practice with writing mechanics within students' own writing", "To gain practice with particular forms of writing (e.g., letter writing)", and "To gain practice with forms of writing encountered on standardized tests". Teachers rated on a six-point scale: 0= "Never", 1= "Fewer than 5 times", 2= "5 times or more", 3= "Monthly", 4= "Weekly", and 5= "Daily". I aggregated these items into one composite variable by taking the mean because these items describe the same latent trait of writing purposes (α =0.91).

The engagement of students in writing processes in Year 3: In the 2009-10 survey, teachers were asked to rate how often they had students engage in several writing-related activities on a six-point scale: 0= "Never", 1= "Fewer than 5 times", 2= "5 times or more", 3= "Monthly", 4= "Weekly", and 5= "Daily". These activities included "Brainstorming or organizing ideas for writing text", "Composing text", " Revising text (focused on meaning and ideas)", "Editing text (focused on grammar, usage, punctuation, spelling)", "Meeting individually with the teacher to get oral feedback or discuss how to improve his or her writing",

"Reviewing written feedback on their own writing given by the teacher", "Sharing or presenting their own writing to peers", and "Analyzing what makes particular texts good or poor models of writing (individually or with others)". I aggregated one composite variable by averaging the ratings on these items (α =0.96).

Focal Predictors

PD Duration in Year 3 (2009-10 School Year): In our 2009-10 survey, researchers asked teachers to indicate how many hours of professional development related to teaching writing or assessing writing they had participated in as a recipient, including workshops, conferences, classes, writing groups, and site-based professional development activities such as study groups or work on writing with a literacy coach or mentor.

PD Content in Year 3 (2009-10 School Year): In the spring 2010 survey, teachers were asked to indicate the extent to which their PD in writing had focused on 12 aspects of writing instruction-related knowledge and strategies on a three-point scale: 0= "not a focus", 1= "minor focus", 2= "major focus". I then aggregated a composite variable by taking the mean of eight items based on factor analysis ($\alpha=0.87$). These eight items measured one common construct of the content knowledge necessary for teaching writing, and they included "Improving student skills and knowledge of planning and pre-writing strategies (brainstorming, generating and organizing ideas, identifying purpose and audience)," "Improving student skills in drafting, revising, and editing text (for meaning, clarity, sentence structure, word choice)," "Improving student skills in grammar, usage, punctuation, or spelling, " "Improving student ability to work collaboratively with their peers on writing, " "Improving student skills for analyzing models of good writing and applying insights to their own text," "Improving student learning about literary techniques and authors' styles", "Improving collaboration among teachers on writing instruction

(either within a single subject or grade level or across the curriculum)," and "Learning about writing by writing yourself and revising your own work with other teachers".

PD format of active learning strategies in Year 3 (2009-10 School Year): To create a measure of PD format, I aggregated one composite variable by taking the sum of 15 items $(\alpha=0.88)$ that describe activities that teachers had participated in as part of any writing PD during the 2009-10 school year. These 15 items include "I received coaching or mentoring in the classroom," "I met formally with other participants to discuss classroom implementation," "I practiced under simulated conditions and received feedback," "My teaching was observed by the professional development provider(s) and feedback was provided," "My teaching was observed by other participants and feedback was provided," "I communicated with the professional development provider(s) concerning classroom implementation," "My students' work was reviewed by participants or the professional development provider(s)," "I met informally with other participants to discuss classroom implementation," I developed curricula or lesson plans that were reviewed by other participants or the professional development provider(s)," "I gave a lecture or presentation to colleagues or other participants," "I conducted a demonstration of a lesson, unit or skill," "I led a whole-group discussion with colleagues or other participants," "I led a small-group discussion with colleagues or other participants," "I wrote some text (e.g., a reflection, plan, poem, etc.)," and "I created rubrics or used rubrics to assess student work."

Analytic Strategies

This study analyzes general PD effects, which may include National Writing Project related PD, but are not limited to such PD. To isolate PD effects from the treatment effects, I conducted separate analyses within each experimental condition, that is, within the treatment or the control group (Nye et al., 2004). In what follows, I introduce some analytical strategies to
establish the confidence of drawing causal inferences between effective PD features and the two outcomes of interest.

Estimation of the Number of Colleagues Helped with Teaching Writing

The logic of estimation is straightforward; that is, I assume that the change in the number of colleagues a teacher helps is a function of the PD experienced by the teacher. However, potential challenges to causal inference demand strategic approaches to eliminate alternative paths from PD to the change in the extent to which teachers committed to helping others. Even if it can be verified that PD features (the causes) have strong associations with the number of other colleagues helped (the outcome), I still cannot confidently claim the causal relationship because there might be other variables that drive either the outcome or causes, or both.

Although there were no average differences in baseline school characteristics between treatment and control schools, because there might be differences among schools within each condition I controlled for school fixed effects. To statistically account for the differences in pretreatment characteristics of individuals and to establish a condition under which causal estimates are comparable to those produced under the randomization at individual levels, I controlled for the prior number of colleagues helped in the 2008-09 school year. In fact, recently, Cook, Shadish and Wong (2008) and Shadish, Clark, and Steiner (2008) showed that estimates from non-randomized studies that included controls for the precondition of the outcome variable closely approximated estimates from randomized experiments. First, teachers retained their behaviors from prior time points, and thus the prior number of colleagues helped would correlate strongly with the outcome. Controlling for the prior could substantially reduce the amount of predictable errors and thus increase the precision of estimation. Second, the prior absorbed the influence of other unmeasured and sustaining characteristics of teachers, such as personal value

placed on collaboration (Frank, 2000). Controlling for the prior potentially reduced the bias due to the impact of these characteristics on the outcome variables and on the amount of PD received.

Except for controlling for the prior number of colleagues helped, I also accounted for instructional practices in Year 2. Instructional practices were used as a proxy of teachers' expertise. The teacher became a "go-to" expert because of having such knowledge and resources that other colleagues who wanted to be effective might desire. Although teaching writing was a cross-subject activity, compared to teachers in other subject areas, teachers who taught English Language Arts (ELA) might be more likely to provide help with writing because they were content experts. I thus included a dummy variable indicating whether the teacher was an ELA teacher in Year 3. Having a degree of Master's or higher was used as another proxy of teachers' knowledge. If a teacher had a Master's degree or higher in Year 3, the dummy variable assigned her/him as "1", otherwise, "0". The last proxy of teachers' expertise was their role in the school. If teachers were instructional coaches and/or teacher consultants, they were expected to be more involved in PD and more likely to provide help to colleagues.

I also controlled for teachers' working experience at the current school, because the relationship between experience, expertise and number of others helped can be mixed. On the one hand, teachers with more experience accumulate more subject and pedagogical knowledge and skills from trials and errors of teaching mathematics. The longer the teacher had been teaching in the school, the more the teacher would acquire local knowledge about the school, the community, and the students. This local knowledge might be desired by other novice teachers and, therefore, might affect the number of teachers who had sought help from them. On the other hand, if experiences make teachers less contingent and flexible for different groups of students, or redundant to respond to new instructional expectations, these experiences actually stymie

gains in expertise (Borko & Livingston, 1989; Reynolds, 1992) and make them less attractive to other teachers. To account for possible effects of teaching experience, in the spring 2010 survey, the WPPD researchers asked teachers to fill in the total number of years they had been teaching in this school.

Finally, the pressure on the school to improve student performance on the state writing assessment might motivate teachers to collaborate or improve their own effectiveness, because the collaborative incentives may be attached to the performance-based accountability (Kelley & Protsik, 1997), or because improving the outcome of teaching writing is a collective activity (Cobb et al., 2003). To control for this contextual factor, in the spring 2010 survey, teachers were asked to rate the amount of pressure that they perceived related to state writing assessment on an eight-point scale (0-7). The estimation model is simplified as:

The number of colleagues helped in Year $3_i = \beta_0 + \beta_1 PD$ duration in Year 3_i

(2.1)

- Or + $\beta_1 PD$ content in Year β_i
- Or + β_1 *PD format in Year 3*_{*i*}
- + β_2 The number of colleagues helped in Year 2_i
- + β_3 Instructional practices in Year 2_i
- + β_4 Being an ELA teacher in Year β_i
- + β_5 Being a female *i*
- $+\beta_6$ Years of working at the current school up to Year 3_i
- + β_7 Being a coach or teacher consultant in Year 3_i
- + β_8 Having a master's degree or higher in Year 3_i

 $+ \beta_9$ Perceived pressure on improving student performance on state writing assess in Year 3_i

+ $\sum \beta_p$ School dummy variable *i*+ e*i*

Where $\beta_{1.9}$ is the coefficient of each predictor, which represents the direction and strength of association between each predictor and the outcome variable. β_p represents the school fixed effect where teacher *i* worked. There are 19 school fixed effects in the modeling of treatment condition and 18 in the control condition. e_i is assumed to be normally distributed with mean 0 and variance of σ^2 .

Estimation of How PD Shapes Instructional Practices through Collegial Interactions

I first estimated the contribution of Year-2 PD features to instructional practices in spring 2009. Then I used techniques of social network analysis to examine the extent to which the learned Year-2 PD was distributed by PD participants to other teachers through professional interactions during year 3, and thus changed other teachers' instructional practices in spring 2010.

To estimate the amount of expertise learned from PD, I used teachers' reported PD features in Year 2 to predict teachers' instructional practices in Year 2, by controlling for Year-1 instructional practices. Then I multiplied the coefficients of Year-2 PD features with the observed values of PD features to get the estimate of the amount of changes in instructional practices attributable to received Year-2 PD. About 50% to 60% of the total variance of Year-2 instructional practices has been explained by these models. The coefficients of PD features are listed in Table 2.3^2 , which are positively significant at the 1 percent level (p-values<0.001).

Insert Table 2.3 in Appendix 2.B about Here

² I examined the impact of other factors that might reduce or invalidate PD effects on instructional practices in year 2. By including all possible confounds, the R-squares of the estimation model did not increase significantly and the coefficients of PD features did not vary significantly. Therefore, the estimates of PD coefficients in table 3 are relatively robust to these alternative model specifications.

To illustrate the dynamics of how expertise was diffused among teachers, I developed a measure of the exposure of a teacher to colleagues' PD expertise through her direct interactions. To measure teachers' interactions, in the spring 2010 teacher survey, teachers were asked to list five colleagues in the same school who had provided help with teaching writing to them during the 2009-10 school year. Teachers were also asked to rate the frequency of each of the five types of interactions on a five-point scale "0=not at all," "1=once or twice this year," "2=monthly," "3=weekly," and "4=daily," including "Gave me curriculum resources (e.g., texts, lesson plans, print materials for students)," "Gave a demonstration of how to lead a writing lesson or activity," "Provided me with feedback on my teaching that I used to improve how I teach writing," "Gave me an idea for a new writing-related activity to use with my students," and "Helped me adapt or improve a writing activity I used with my students." The original units of the frequency of interactions were transformed to days (0=0 days, 1=2 days, 2=10 days, 3=36 days, 4=180 days). I then summed the frequency of interactions between two teachers across these different types of interactions. For instance, teacher Lisa nominated Bob as a help provider. Bob had given Lisa curriculum resources monthly (10), a demonstration of instruction once or twice in this year (2), and an idea of new writing-related activity every week (36). Thus, given the pair of these two teachers, Lisa and Bob, I would calculate the frequency of their interactions as the sum of these frequencies to be 48 (10+2+36).

The direct exposure to help providers' instructional expertise gained from year-2 PD participation (short as "providers' PD expertise") was approximated by multiplying the frequency of the interaction teacher *i* reported with *i*' by the estimated amount of knowledge that teacher *i*' learned from PD in year 2. For example, if Bob's PD expertise was 2 and the

frequency of Lisa and Bob's interaction was 48, then Lisa's exposure (via Bob) would be 48 x 2 = 96. If besides Bob, Lisa also nominated Lucy with PD expertise of 2 (at a frequency of 180, then 180 x 2 = 360), Tracy with PD expertise of 0.1 (14, 14 x 0.1=1.4), and Tom with PD expertise of 5 (10, 10 x 5=50). To combine information across Lisa's network, I took the sum exposure across all teachers that Lisa nominated between 2009 and 2010:

Direct exposure
$$_{i} = \sum_{\substack{i'=1, \ i\neq i'}}^{n_{i}} (Help_{ii'}) \times (Providers'PD expertise_{i'})$$
 (2.2)

Where in equation (1) n_i is the number of teachers *i* (e.g., Lisa) indicated as providing help with writing instruction (e.g. $n_i = 4$) and help_{*ii*}, represents the frequency with which teacher *i* (e.g., Lisa) reported receiving help from *i*' (e.g., Bob). In previous example, the direct exposure of Lisa to her colleagues would equal 507.4 (96+360+1.4+50).

Teachers' instructional practices in Year 3 were then examined as functions of colleagues' direct exposure to peer's PD expertise through interactions (Frank & Fahrbach, 1999) after accounting for individuals' practices in Year 1, town participation in professional development in Year 3, and personal background characteristics in Year 3, as well as school fixed effects. The model is simplified as

Instructional practices in Year $3_i = \beta_0 + \beta_1$ Own experienced PD features in Year $3_i + \beta_2$ Direct exposure to PD features experienced by one's peers in Year 3_i

- + β_3 Prior instructional practices in Year I_i
- + β_4 Being an ELA teacher in Year β_i
- + β_5 Being a female *i*
- + β_6 Years of working at the current school up to Year 3_i (2.3)

+ β_7 Being a coach or teacher consultant in Year β_i

+ β_8 Having a master's degree or higher in Year 3_i

+ β_9 Perceived pressure on improving student performance on state writing assess in Year 3_i

 $+\sum \beta_p$ School dummy variable $_i$ + e_i

Moreover, given the strong correlation among three measures of PD features according to Cohen³ (ρ >0.3, p-value \leq 0.001), as indicated in Table 2.4, I added them separately into the model to avoid multicollinearity issues.

Insert Table 2.4 in Appendix 2.B about Here

Quantifying Robustness of Inferences

Any policy or theoretical interpretations I make in this study will depend on the robustness of inferences. Recognizing the importance of causal inference and no matter how many statistical controls that I employ, there will be inevitable concerns about the validity of inferences. Therefore to inform discourse about inferences, I quantify the concerns about the potential to invalidate these inferences. This approach can be considered an extension of sensitivity analysis (e.g., Copas & Li 1997; Robins, Rotnitzky, &Scharfstein 2000; Rosenbaum & Rubin 1983).

³ Cohen gives the following guideline for evaluating the strength of the relationship by using Pearson product correlation in social sciences: small effect size, r = 0.1 - 0.23; medium, r = 0.24 - 0.36; large, r = 0.37 or larger. Moreover, the sample sizes in this study are between 850 and 170, which indicate even stronger associations among these three measures of PD features. Refer to: Jacob Cohen (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.

Cohen, J (1992). A power primer. Psychological Bulletin 112: 155–159.

Classically, internal validity can be expressed in terms of confounding variables that are correlated with both the predictor of interest and the outcome (Shadish, Cook, & Campbell 2002). For example, the effects of PD features could be confounded with Year-3 motivation to attend professional development because motivation could be correlated with the type of professional development received as well as subsequent changes in practices. This is also known as concern over selection bias (Heckman, 1978) or identification (Manski, 1995).

To express robustness that accounts for the relationship between a confounding variable and the predictor of interest *and* between the confounding variable and the outcome, Frank (2000) defines the impact of a confounding variable on an estimated regression coefficient as impact= $r_{yv} \times r_{xv}$. In this expression, r_{yv} is the correlation between a confounding variable, v(e.g., motivation), and the outcome y (e.g., change in teaching writing), and r_{xv} is the correlation between v and x, a predictor of interest (e.g., PD features). Frank (2000) then quantifies how large the impact must be to invalidate an inference.

Results

Descriptive Statistics of PD features in Both Partnership (Treatment) and Delayed Partnership (Control) Schools

Table 2.5 indicates significant mean differences in these three PD features between partnership schools and delayed partnership schools. Teachers in partnership schools, on average, participated in three times as many hours of PD as peers in delayed partnership schools. Also, teachers in partnership schools participated in PD with high-quality content and format twice as often as peers in delayed partnership schools. The clear treatment effects on PD features (the causes) make it necessary to use stratification (conducting separate analysis in treatment and

control condition) to account for these treatment effects on the causal inference of the relationship between PD features and the outcomes of interest.

Insert Table 2.5 in Appendix 2.B about Here

Effects of PD Features on the Number of Colleagues Helped With Teaching Writing

Estimated Effects: Table 2.6 shows the estimated effects of PD features on the number of others helped with teaching writing from six models, separately for each PD feature and each experimental condition. Results from Model-I of estimating the effect of PD duration are included in the second column; results from Model-II of estimating the effect of PD content are included in the third column; and results from Model-III of estimating the effect of PD format are contained in the fourth column. Overall, each of the three models explains about 50%-60% of the total variance of the number of colleagues helped during 2009-10 school year.

Insert Table 2.6 in Appendix 2.B about Here

The t-ratio of the predictor of PD duration on the number of colleagues helped is 2.06 with an effect size (Cohen's d) of about 0.25, which can be referred to as a small effect (Cohen, 1988)⁴. The unstandardized coefficient is 0.012 in the treatment condition and 0.028 in the control condition. Translating this statistic to a real-life example of a school with 10 teachers who could have had 20 hours more of PD out of 50 teachers in a school, there would have been

⁴ Cohen labeled an effect size small if Cohen's d = .20 or correlation coefficient r = .10, large if d = .80 or r = .50, and medium if d = .50 or r = .30. Cohen, J. *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum

possibly 2 additional teachers in treatment condition and 5 additional teachers in the control condition being helped. This is a discernable piece of evidence on PD participants' spillover effects.

The unstandardized coefficient of PD content on the number of colleagues helped is 0.695 in the partnership condition (Cohen's d = 0.26) but close to zero in the delayed-partnership condition. The difference might be caused by the treatment effect; that is, PD programs offered to teachers in partnership schools emphasized more on this content than those offered to peers in control schools. Peers in delayed-partnership schools were exposed to the same PD features, but the dosage was not strong enough to have an effect.

The variable of PD format is a significant predictor of the outcome variable in both conditions. The estimate of the effect of PD forms on the number of colleagues helped is 0.23 in the partnership condition and 0.13 in the delayed partnership. T-ratios suggest a large effect size (Cohen's d) of 0.75 in treatment (partnership) condition and medium effect size of 0.36 in control (delayed partnership) condition. Extrapolating to a school where 10 teachers out of 50 teachers experienced one additional PD format of engaging teachers in active learning, additional 2 teachers in the treatment group or one teacher in the control group would have benefited.

Moreover, the strongest predictor of the number of colleague help in year 3 is the prior number of colleagues helped in year 2. Its unstandardized coefficient is 0.5 or larger (p-value <0.001) and it explains one half of the variance of the outcome variable. Moreover, help with teaching writing is more likely to be sought from ELA (English Language Arts) teachers than teachers who taught other subject areas, which is predictable because ELA teachers are likely to be respected as content experts with regard to teaching writing. I then notice that the standardized coefficients of PD features (duration, content and format) are similar to that of

being an ELA teacher, which indicates that the effects of PD features are comparable to that of being an ELA teacher.

None of the rest of the covariates, such as teaching experience, being a coach or teacher consultant, being a female, perceived pressure, or having a Master's degree or higher, significantly predicted teachers' helping with others.

Quantifying Robustness: Using Frank's calculation, the impact of an unmeasured confound would have to be greater than 0.153 to invalidate the inference. In terms of correlation components, r_{yy} (the correlation between the confound and the number of colleagues helped) must be greater than 0.368 and r_{xy} (the correlation between the confound and PD format) must be greater than 0.415 to invalidate the inference (using Frank's multivariate correction), which are strong correlation in social science according to Cohen's criteria⁵. Making an interpretation of this correction more intuitive, it is helpful to compare the threshold to the impacts of measured covariates. By partialling out the prior number of colleagues helped, being an ELA teacher has the strongest impact among measured covariates. Its impact on the inference of PD format on the number of colleagues helped in year 3 is 0.03, which is the product of the partial correlation with PD format (0.336) and the partial correlation with number of colleagues helped in year 3 (0.089). Thus, by controlling for prior number of colleagues helped, the impact of an unmeasured confound necessary to invalidate the inference of PD format would have to be five times stronger than the impact of the strongest predictor of being an ELA teacher; such an unmeasured confound would unlikely exist in practices.

⁵ Cohen's benchmarks are generic descriptors of the magnitude of effect size. Studies in education are likely to smaller effect sizes than other areas (Valentine & Cooper, 2003). Therefore, using Cohen's labels may be misleading and needs to be cautious.

Similarly, to invalidate the inference of PD duration on the number of colleagues helped, the impact of an unmeasured confound would have to be greater than 0.027 and r_{yv} (the correlation between the confound and the number of colleagues helped) must be greater than 0.155 and r_{xv} (the correlation between the confound and PD duration) must be greater than 0.176. These are medium strength of associations. That is, the inferences on PD duration could be violated by a variable that has medium correlation with both PD duration and the number of colleagues helped in year 3 and has an impact that is almost as strong as the impact of the covariate of being an ELA teacher (0.03). This implies a medium-level robustness of inference.

Lastly, the impact of an unmeasured confound must be greater than 0.005 to invalidate the inference of PD content on the number of colleagues helped in year 3. Correspondingly, r_{yv} must be greater than 0.07 and r_{xv} must be greater than 0.08 to invalidate the inference. These are low strength of associations and imply low-level of robustness of inference.

Effects of PD Features on Instructional Practices through Professional Help

Estimated Effects: First, I controlled for the effect of a teacher's own PD when estimating the influence of exposure to peers' PD. Second, I examined two types of instructional practices, the breadth of writing purposes taught in year 3 denoted as model-I in Table 2.7, Table 2.8, and Table 2.9 and the engagement of students in writing processes, denoted as Model-II.

Consistent with some previous studies, PD duration has a significantly positive impact on each of these two types of instructional practices, for teachers in both treatment and control schools, as shown in Table 2.7. The effect sizes vary between 0.23 (β =0.005, b=0.083, Cohen's d=0.23) and 0.4 (β =0.024, b=0.186, Cohen's d). The effect size in the treatment condition is lower than that in the control condition on both of these two measures of instructional practice. Moreover, after controlling for teachers' own PD contact hours, their prior practices in year 1,

and other covariates, interactions with peers who were involved in intense PD in year 2 would have a significantly positive impact on these teachers' instructional practices in Year 3 in both treatment and control schools. The effect of peers' influence is estimated to be between β =0.13 (b=0.077, Cohen's *d*=0.26) and β =0.144 (b=0.136, Cohen's *d*=0.38).

Insert Table 2.7 in Appendix 2.B about Here

As shown in Table 2.8, although the impact of PD content on the breadth of writing purposes taught by teachers in the control group was not statistically significant, overall, the results suggest a strong and positive impact of teachers' reported PD content on instructional practices (β ranges from 0.461 to 0.457; b ranges from 0.145 to 0.222; t-ratio ranges from 3.32 to 4.51, Cohen's *d* ranges from 0.33 to 0.43). Exposure to peers' experienced PD content has positive effects too as included in the fourth row of Table 2.8. With a one standard-deviation increase in exposure to the PD content experienced by peers, teachers would have positively improved their instructional practices by about 0.1 standard deviations.

Insert Table 2.8 in Appendix 2.B about Here

As shown in Table 2.9, teachers who had participated in PD with various PD formats that engaged teachers in active learning had a higher likelihood of improving their writing practice in both the partnership and the delayed partnership groups. The effect of PD format on teachers' engagement of students in writing processes in the treatment group had the largest effect size. Its unstandardized coefficient β equals to 0.082 and standardized coefficient *b* equals to 0.197, and t-ratio equals to 5.24 (Cohen'd=0.5), a medium effect according to Cohen's benchmark. After controlling for all predictors including one's own PD and prior practices in year 1, exposure to PD formats experienced by one's peers could change one's instructional practices. The coefficients shown in the fourth row of Table 2.9 indicate the hypothesized positive effect and relatively substantial magnitudes of spillover effects of PD participants who had experienced these high-quality PD activities.

When comparing the standardized coefficients of exposure to peer's PD expertise to those of teachers' own PD features in Tables 2.7, 2.8 and 2.9, these peer effects are close to those of own experienced PD, which deserves our attention.

Insert Table 2.9 in Appendix 2.B about Here

Quantifying Robustness: I only quantify the inference with regarding to PD participants' spillover effects. The impact of an unmeasured confound would have to be greater than 0.067 to invalidate the inference of peers' Year-2 PD duration on teachers' own practices of engaging students in writing processes (β =0.144). Correspondingly, a confounding variable would have to be correlated with the engagement of students in writing processes at 0.227 and with PD duration at 0.296, which are medium correlations. Comparing this impact to the impact of the measured covariate of teacher's own PD duration in Year 3, the impact of teacher's own PD duration in Year 3 is about 0.012, the product of the correlation with exposure to peers' PD (ρ =0.084) and the correlation with the engagement of students in writing processes in Year 3 (ρ =0.148). The impact of an unmeasured confound necessary to invalidate the inference of peers' spillover effect

would have to be five or six times stronger than the impact of teachers' own PD duration. This unmeasured confound may rarely exist in practice.

Similarly, the impact of an unmeasured confound would have to be greater than 0.051 to invalidate the inference of exposure to peers' expertise gained from receiving Year-2 PD content on the engagement of students in writing processes in year 3 (β =0.128), while r_{yy} (the correlation between the confound and the engagement of students in writing processes) must be greater than 0.2 and r_{xy} (the correlation between the confound and the exposure to peers' year-2 PD content) must be greater than 0.254 to invalidate the inference. These component correlation coefficients imply a medium-level robustness of inference. When comparing to the strongest observed covariate, the impact of an unmeasured confound necessary to invalidate the influence of PD content experienced by one's peers must be stronger than the impact of one's own PD content (0.046, the product of the correlation with the engagement of students in the writing process in year 3 (ρ =0.29) and the correlation with peers' influence (ρ =0.16)).

Finally, the impact of an unmeasured confound must be greater than 0.055 to invalidiate the inference of exposure to PD formats experienced by peers on teachers' breadth of writing purposes taught in Year 3, and r_{yy} (the correlation between the confound and the breadth of writing purposes taught) must be greater than 0.203 and r_{xy} (the correlation between the confound and the exposure to peers' year-2 PD formats) must be greater than 0.27 to invalidate the inference. The impact of an unmeasured confound necessary to invalidate the influence of PD format experienced by one's peers would have to be twice as strong as the impact of one's own experienced PD format (0.03, the product of the correlation with the breadth of writing purposes taught in year 3(ρ =0.17) and the correlation with the exposure to peers' year-2 PD formats (ρ =0.179)).

Discussion

This study investigates two questions related to how PD characteristics can promote knowledge diffusion and writing instructional improvement. In particular, I examined two aspects of the spillover effect of PD: on the number of colleagues helped and on peer influence on instructional practices through helping. After analyzing longitudinal data from an experimental evaluation study, I found that teachers were more likely to provide help to others with teaching writing if they intensively participated in high-quality PD programs. The PD duration (contact hours), content (substance), and format (delivery methods) had small, medium, or large effects on the number of colleagues helped in year-3 , after accounting for prior number of colleagues helped and other important confounds. Moreover, I found that the expertise that teachers gained from year-2 PD spread out to other teachers as they offered professional help. In some cases, the peer effects on the improvement of instructional practices are almost equal to the effects of teachers' own PD. In this section, I further interpret these findings in terms of their theoretical, methodological, and policy implications.

Substantive Interpretation

This study reveals important effects of several PD features (i.e., duration, content and format). For example, teachers that were exposed to longer contact hours were more likely to improve their instructional practices. Moreover, teachers were more likely to change their teaching practices if the foci of the writing PD placed greater emphasis on: a) strategies for improving students' writing skills and knowledge, and students' ability to work collaboratively with their peers, b) teachers' own writing skills and their ability to collaborate with colleagues in developing teaching strategies. Additionally, teachers benefited from participating in PD activities if they a) received in-classroom coaching or mentoring, b) actively discussed classroom

implementation with co-participants or PD providers, c) analyzed students' work with other teachers, and d) received constructive feedback on their classroom teaching. These effective PD features have been discussed in other studies. However, this study distinguishes itself from previous research and contributes to the literature in several ways.

First, this study provides more robust estimates of the effects of high-quality PD features on instructional practices. As noted in previous sections, Garet et al. (2001) illustrated the positive effects of PD features in promoting teachers' knowledge and changes in classroom practices by using a cross-sectional design with one year of data. Although they provided the first large-scale analysis of PD features, due to the cross-sectional design, they were not able to rule out other alternative explanations for the association between PD features and teachers' knowledge and practices. Desimone et al. (2002) investigated the same effective PD features and drew on the same sample of teachers as Garet et al. (2001). Unlike Garet et al.'s study, they used data over three years and estimated Year-2 PD effects on Year -3 instructional practices, by controlling for Year-1 instructional practices. This analytical strategy was more helpful to draw causal inference between PD features and change in instructional practices. However, Desimone et al.'s study was limited by a small sample size⁶, as the authors pointed out in the paper. Their study might not have enough analytical power to carry out the complex longitudinal analysis as intended. To complement the above two studies, I used data from a larger sample of teachers. More importantly, the use of several statistical techniques for estimating causal inferences which will be reviewed in the next section leads to more robust estimates of PD effects.

Second, this study untangles the mechanism of how intra-organizational dynamics can mediate the PD effect on instructional practices, which is far from well understood in the

⁶ The sample size at teacher level varied from100 to 130 across models.

literature. Using sociometric data, I am able to identify with whom teachers had interacted, the frequency of their interactions, and the substance of their interactions, which comprise the direct measure of peer interactions. I thus was able to demonstrate the extent to which effective PD programs could shape professional interactions among teachers and the extent to which teachers' professional interactions could augment the impact of PD on instructional practices. These PD spillover effects are estimated to be close to the direct effects of PD on instructional change. Such substantial effects, however, have been long-time ignored in prior estimations of PD impact on teaching and learning outcomes.

Moreover, this study probes into the "black box" of program implementation within schools and helps understand how school intra-organizational mechanisms can mediate the impact of external intervention on teaching and learning outcomes. The diffusion of external intervention within schools is not linear; rather, it is similar to the fact that "light bends in the water". Even the degrees of "refraction" differ when light passes through different surfaces. Similarly, the implementation of the same educational intervention (light) may vary across schools (surfaces) because school contextual mechanisms may change the direction and the magnitude of the impact of external interventions. To better understand how interventions may work in schools, we need to develop explanatory and predictive theories of these school processes and mechanisms.

Third, I estimated the PD effects on classroom practices independently from peer learning. Although it has been long acknowledged that teachers' immediate social context (i.e., teachers' professional networks) enables or constrains their behaviors and beliefs, it is hard for prior studies to control for teachers' learning from peers when estimating the amount of improvement in knowledge and skills that could be attributable to learning in PD. This study fills

the gap in the literature and examines the effects of teachers' own PD by controlling for peer influence, and vice versa.

Methodological Interpretation

Using Cohen's benchmarks, some of the PD effects identified in this study may not be large enough to attract policy makers' attention. Also, using Frank's robustness index, some of the PD effects are not as persuasive as expected (for example, the effect of PD content on the number of colleagues helped in year 3). However, I still have relatively high confidence in establishing these causal inferences between PD features and outcomes of interest because I used multiple strategies to eliminate the impacts of measured and unmeasured confounds on regression estimates. The significance of each strategy in terms of drawing causal inference is further reviewed as follows:

First, the t-test results in Table 2.1 and Table 2.2 showed that there were no significant mean differences in several observed school contextual factors, which indicated the success of implementing the randomization at the school level. I then capitalized on this randomization to examine the extent to which PD could affect writing instructional practices within each experimental condition. This strategy separated the treatment effects of National Writing Project from general PD effects (Nye et al., 2004). Moreover, I used school fixed effects to control for unobserved heterogeneity across schools.

Second, to account for the substantial variation in pretreatment characteristics of individuals within schools, I controlled for prior conditions of these two dependent variables—the number of colleagues helped and teaching practices. Accounting for the prior made at least three contributions to drawing causal inferences. First, the prior of the outcome variable absorbed the impact of pre-conditional covariates on the outcome variable of interest (Frank,

2000). Including the prior as a covariate in the estimation of the outcome variable helped to create a counterfactual condition, which removed the effect of previous differences in the outcome and assumed all individuals had the same prior to begin with. Second, controlling for the prior increased the precision or power of estimates by reducing the amount of predictable errors, because the prior is usually the most important predictor of the outcome variable. For instance, in this study, the prior explained more than 50% of the total variance of dependent variables. Third, for estimation purposes, controlling for the prior reduced the potential of non-normality (Raykov & Marcoulides, 2008) and thus reduced the potential bias due to the violation of normality assumptions in estimation. Besides controlling for the prior, I also accounted for individual unique characteristics that might confound the influence of PD features.

Last but not least, I used Frank's robustness index to quantify the sensitivity of inferences to unknown, or at least, unmeasured confounds. Given that the purpose of this study is to develop recommendations for educational policies, discussing the robustness of these PD estimates helps policymakers and researchers discern the internal validity of the inferences in this study and the risk of applying these findings to practices. Results shown that except that the inference of PD content on the number of colleagues helped in year 3 had relatively low-level of robustness, the rest of inferences of PD spillover effects had medium-level or strong-level robustness to any unmeasured omitted confounding variables. That is, we should have substantial confidence in the internal validity of identified PD spillover effects.

Limitations

This study has some obvious limitations. First, this study lacks information about students. I was not able to rule out a possible cohort effect; that is, differences in characteristics of student cohorts from Year 1 to Year 3 might affect the ways in which teachers taught.

However, the student cohort effect would have only invalidated the inference of PD effects when the changes in student composition were correlated with these PD features under investigation. Second, I only examined three, but not all, PD features. It is possible that these unexamined PD features drive the positive spillover effects identified in this study. Future studies can further explore and provide empirical evidence on these possibilities.

Third, the findings about the dynamics of teacher interactions uncovered in this study need to be enriched and confirmed by qualitative evidence. We need to know more in qualitative terms about the collaborative dynamics. Fourth, the larger WPPD project aimed to evaluate the partnership between schools and Local Writing Project sites. Schools were the units of analysis and thus the randomization happened at school level. However, for this study of estimating the impact of PD features on the change of teachers' instructional practices, teachers become the units of analysis, and it would be ideal if teachers were randomly assigned to receive different levels of PD features. Although statistical control might be able to provide sufficiently robust evidence to support these results, conducting experimental studies to further verify or falsify these results in the future is still valuable before policymakers put them on their agenda.

Policy Implications

Some policy implications stem from the findings of this study. First, this study provides evidence to support the assumption that PD is a vital tool to improve teaching quality. Some teachers may be born to teach, but most teachers learn to teach. If we presume that all students can be educated, we should also believe that teachers can be educated and improved, too. Rather than basing policy primarily upon attracting good teachers or firing bad ones (c.f. Hanushek, 2009), we can think of how to use in-service PD to effectively develop teachers that we desire, keep them updated, and prepare them ready for classroom problems. Furthermore, to allow the

incentive strategies (such as merit pay) to have an effect on improving teacher quality, we need to provide the avenues and resources for teachers to learn and grow. More importantly, this study examined another aspect of PD effects, the participants' spillover effect. This indirect effect has been ignored in prior calculation of PD effects. But if collaboration and learning from peers is important to improve the performance of all school faculty, then PD effects for promoting the provision of help should be relevant to policymakers, too.

Second, there might be several challenges to implementing these advocated PD features in practice. The promoted PD features in this study include personalized coaching and mentoring and collaborative activities, which are more costly than traditional workshops and conferences. Moreover, teachers might struggle to find the time to meet and find a way to establish effective team norms. Further studies (such as cost-effectiveness analysis) have to be done to explore effective strategies to cope with these problems.

Third, legitimating some PD receivers' roles as teacher leaders would be a way to wisely use PD spillovers. It is valuable to think about how to distribute PD receivers in the organizational structure to shape the knowledge diffusion among teachers. For instance, grouping PD receivers with low-performing colleagues or non-PD participants on the same task may promote the knowledge flow from PD receivers to these peers.

Conclusion

Beyond exploring the direct effect of PD on participants' own instructional practices, this study fills a gap in literature by examining the spillover mechanisms of PD participants on other colleagues' instructional change through intra-school networks. PD participants become more likely to provide help to other peers and through the provision of help, they could improve other colleagues' instructional practices. The inferences have been consolidated by estimation

strategies used in this study aiming to establish confidence of drawing causal inference. Built on such confidence, I propose several policy suggestions in terms of investment in these effective PD features and distribution of PD receivers within schools. However, there were limitations of this study in research design and data analysis. If the limitations were important enough to invalidate these causal inferences, policy suggestions given in this study would be irrelevant. Thus, in the end, I call for future studies to examine the findings of this study in other educational contexts or to use other research designs, such as random assignment of teachers or qualitative data analysis. **APPENDICES**

APPENDIX 2.A

Figure 2.1 The Theoretical Framework of How Features of PD Affect the Dynamic of Diffusing Expertise among Teachers



APPENDIX 2.B

	Partnership	Delayed Partnership
	(Treatment)	(Control)
Mean enrollment	669.29	564.84
Wear enronment	(368.14)	(268.58)
Mean % FDD	44	53
	(25)	(26)
Moor % White	64	58
Mean % white	(28)	(30)
Maan munil too shan notio	15.37	14.16
Mean pupil-teacher ratio	(2.96)	(2.98)
Mean full time equivalent (FTE)	46.93	42.29
teachers	(24.4)	(23.79)
Mean 7/8 English language arts	4.63	4.18
(ELAs)	(3)	(3.07)

Table 2.1 School Characteristics in 2008-09

Note: Standard deviations appear in parentheses.

	Partnership	Delayed Partnership
	(Treatment)	(Control)
Experience		
Mean years teaching	13.56 (SD=9.87)	12.97 (SD=9.52)
Mean years teaching in the current school	8.82 (SD=7.94)	7.88 (SD=7.41)
Mean years teaching the same assignment in the current school	7.37 (SD=7.17)	6.67 (SD=6.65)
Highest Academic Degree		
Percent with Bachelor's	41.29 (n=346)	43.09 (n=340)
Percent with Master's	51.67 (n=433)	47.91 (n=378)
Percent with Education Specialist's	5.61 (n=47)	5.2 (n=41)
Percent with Doctorate	0 (n=0)	1.27 (n=10)

Table 2.2 Teacher Characteristics in 2008-09 School Year

Note: In parentheses, SD=standard deviation, n=sample size.

PD features	The breadth of	writing purposes	Engagement of students in the				
	tau	ight	writing process				
	R-square	Coefficient	R-square	Coefficient			
PD duration	0.5	0.009*** (0.001)	0.54	0.41*** (0.072)			
PD content	0.5	0.32*** (0.1)	0.56	0.009*** (0.002)			
PD format	0.52	0.1*** (0.018)	0.56	0.1*** (0.019)			

Table 2.3 Estimates of the Contribution of PD Features To Year-2 Instructional Practices

Notes: Standard errors are included in the parentheses. *** p-value ≤ 0.001

Table 2.4 Pearson Correlation between PD Features

	PD duration	PD contents	PD forms
PD duration	1.000		
PD content	0.308***	1.000	
PD format	0.552***	0.435***	1.000

Note: *** p-value ≤0.001

Table 2.5 Descriptive Statistics of PD features in both Partnership (Treatment) and Delayed Partnership (Control) Schools in 2009-10

	Partnership (Treatment)	Delayed Partnership (Control)	T-Value on Mean Difference
PD duration	10.17	3.76	6.61***
	(19.49)	(3.76)	
PD content	0.67	0.3	9.47***
	(0.64)	(0.52)	
PD format	2.83	1.18	8.77***
	(3.23)	(2.41)	

Note: Standard deviations are included in parentheses. *** p-value ≤ 0.001

		(Partner Treatment.	ship n=265	5)		Delayed Partnership (Control, n=260)					
	Mode	el-I	Model-III Model-III			Model-I Model-II			Model-III			
	Unstan dardize d	Stan dardi zed	Unstan dardize d	Stan dard ized	Unstan dardize d	Stan dard ized	Unstand ardized	Stan dard ized	Unstand ardized	Sta nda rdiz ed	Unstan dardize d	Stan dardi zed
PD duration in Year 3	0.012* (0.006)	0.08 2					0.028* (0.011)	0.14 5				
PD content in Year 3			0.695* (0.331)	0.14 5					-0.193 (0.359)	- 0.0 40		
PD format in Year 3					0.234** * (0.039)	0.08 6					0.132* * (0.045)	0.06 5
The number of people helped in Year 2	0.613** * (0.052)	$\begin{array}{c} 0.58\\0\end{array}$	0.614** * (0.051)	0.58 2	0.587** * ((0.049)	0.55 6	0.503** * (0.062)	0.51 1	0.494** * (0.063)	0.5 01	0.507* ** (0.062)	0.51 6
Instructional practices in year 2	-0.044 (0.135)	- 0.02 2	-0.137 (0.138)	- 0.06 5	-0.210 (0.130)	- 0.09 7	0.038 (0.123)	0.02 7	0.088 (0.125)	0.0 60	0.014 (0.123)	0.01 3
Being an ELA teacher in Year 3	0.713* (0.351)	0.12 3	0.582 (0.349)	0.09 9	0.46 (0.333)	0.07 9	0.831** (0.288)	0.18 2	0.81** (0.290)	0.1 78	0.821* * (0.286)	$\begin{array}{c} 0.18\\0\end{array}$
Being a female	-0.238 (0.320)	- 0.03 7	-0.213 (0.314)	- 0.03 4	-0.178 (0.299)	- 0.02 8	0.129 (0.318)	0.02 5	0.121 (0.318)	0.0 24	0.145 (0.314)	0.02 9
Years of working at the current school up to Year 3	-0.017 (0.017)	- 0.04 8	-0.022 (0.017)	- 0.06 2	-0.018 (0.016)	- 0.05 1	-0.008 (0.018)	- 0.03 1	-0.011 (0.017)	- 0.0 41	-0.008 (0.018)	- 0.02 9

Table 2.6 Estimated Effect of PD Features on the Number of Colleagues Helped With Teaching Writing

Table 2.6 (cont'd)

Being a coach/teacher consultant in Year 3	-0.210 (0.477)	0.02 4	-0.076 (0.464)	- 0.01 2	-0.423 (0.448)	- 0.04 5	0.228 (0.361)	0.0 39	0.275 (0.362)	0.04 6	0.24 (0.358)	0.04 1
Having a Master's degree and higher in Year 3	-0.071 (0.303)	- 0.01 3	-0.018 (0.299)	- 0.00 2	0.022 (0.285)	0.00 5	0.114 (0.249)	0.0 32	0.147 (0.250)	0.04 0	0.068 (0.247)	$\begin{array}{c} 0.02\\ 0\end{array}$
Perceived pressure on improving student performance on state writing assessment in Year 3	-0.032 (0.096)	0.01 4	-0.091 (0.095)	- 0.04 6	-0.049 (0.090)	0.02 4	0.017 (0.075)	0.0 14	0.013 (0.075)	0.01 2	0.004 (0.075)	0.00 5

Notes: Standard errors are reported in parentheses. * p-value≤0.05 ; ** p-value≤0.01; *** p-value ≤ 0.001

		Partne (Treat	ership ment)		Delayed Partnership (Control)			
	Model-I (purposes) (n=434)		poses) Model-II (engagement) (n=432)		Model-I (pur (n=400	rposes)	Model-II (engagement) (n=397)	
	Unstanda rdized	Standar dized	Unstandar dized	Standar dized	Unstandar dized	Standar dized	Unstanda rdized	Standar dized
Own experienced PD duration in Year 3	0.005* (0.002)	0.083	0.007** (0.002)	0.102	0.015** (0.005)	0.137	0.024*** (0.006)	0.186
Exposure to PD duration experienced by one's peers	0.098** (0.032)	0.106	0.144*** (0.036)	0.136	0.141** (0.045)	0.096	0.130** (0.049)	0.077
Prior Instructional practices in Year 1	0.492*** (0.042)	0.495	0.458*** (0.040)	0.474	0.522*** (0.042)	0.529	0.465*** (0.040)	0.467
Being an ELA teacher in Year 3	0.274* (0.107)	0.109	0.567*** (0.123)	0.198	0.213* (0.096)	0.085	0.727*** (0.111)	0.251
Being a female	0.039 (0.089)	0.016	-0.002 (0.010)	-0.001	0.027 (0.094)	0.010	-0.132 (0.102)	-0.045
Years of working at the current school up to Year 3	-0.011* (0.005)	-0.074	-0.002 (0.006)	-0.010	-0.002 (0.006)	-0.011	0.005 (0.007)	0.025
Being a coach/teacher consultant in Year 3	0.051 (0.157)	0.012	0.131 (0.175)	0.026	0.140 (0.129)	0.036	0.27 (0.144)	0.060
Having a Master's degree and higher in Year 3	-0.132 (0.088)	-0.056	-0.095 (0.098)	-0.035	0.212* (0.088)	0.089	0.039 (0.097)	0.014
Perceived pressure on improving student performance on state writing assessment in Year 3	0.031 (0.030)	0.038	0.053 (0.032)	0.057	0.015 (0.028)	0.020	0.028 (0.031)	0.033

Table 2.7 Estimated Effects of PD Duration on Instructional Practices

Notes: Standard errors are reported in parentheses. * p-value≤0.05; ** p-value≤0.01; *** p-value ≤ 0.001

Table 2.8 Estimated Effects of PD Content on Instructional Practices

		Partn	ership		Delayed Partnership				
	Model-I (purposes) (N=434)		Model-II (engagement) (n=432)		Model-I (p (n=40	urposes))0)	Model-II (engagement) (n=397)		
	Unstanda rdized	Standa rdized	Unstanda rdized	Standa rdized	Unstanda rdized	Standa rdized	Unstanda rdized	Standa rdized	
Own experienced PD content in Year 3	0.457*** (0.101)	0.222	0.464*** (0.111)	0.198	0.136 (0.125)	0.049	0.461*** (0.139)	0.145	
Exposure to PD content experienced by one's peers experienced	0.041 (0.035)	0.046	0.128*** (0.036)	0.141	0.144*** (0.044)	0.124	0.112* (0.044)	0.095	
Prior Instructional practices in Year 1	0.460*** (0.041)	0.463	0.431*** (0.040)	0.447	0.521*** (0.042)	0.527	0.454*** (0.041)	0.456	
Being an ELA teacher in Year 3	0.256* (0.104)	0.102	0.543*** (0.118)	0.189	0.217* (0.095)	0.086	0.739*** (0.11)	0.256	
Being a female	0.008 (0.088)	0.003	-0.040 (0.097)	-0.014	0.012 (0.093)	0.005	-0.156 (0.101)	-0.053	
Years of working at the current school up to Year 3	-0.011* (0.005)	-0.071	-0.003 (0.006)	-0.017	-0.002 (0.006)	-0.012	0.004 (0.007)	0.023	
Being a coach/teacher consultant in Year 3	0.06 (0.153)	0.013	0.150 (0.169)	0.030	0.158 (0.129)	0.040	0.301* (0.143)	0.066	
Having a Master's degree and higher in Year 3	-0.110 (0.086)	-0.047	-0.078 (-0.095)	-0.029	0.242** (0.088)	0.102	0.062 (0.096)	0.023	
Perceived pressure on improving student performance on state writing assessment in Year 3	0.018 (0.029)	0.022	0.047 (0.032)	0.050	0.013 (0.028)	0.017	0.029 (0.031)	0.034	

Notes: Standard errors are reported in parentheses. * p-value≤0.05; ** p-value≤0.01; *** p-value ≤ 0.001

Table 2.9 Estimated Effects of PD Format on Instructional Practice	es
--	----

		Partn	ership			Delayed F	Partnership	
	Model-I (purposes) (N=434)		Mode (engage (n=43	l-II ment) 32)	Model-I (p (n=4	(Cor ourposes) 00)	Model-II (engagement) (n=397)	
	Unstanda	Standar	Unstandar	Standar	Unstandar	Standar	Unstanda	Standar
Own experienced PD format in Year 3	0.059*** (0.014)	0.162	0.082*** (0.016)	0.197	0.046** (0.017)	0.101	0.071*** (0.020)	0.135
Exposure to PD format experienced by one's peers	0.065 (0.035)	0.076	0.121** (0.038)	0.127	0.158*** (0.044)	0.134	0.092* (0.048)	0.070
Prior Instructional practices in Year 1	0.465*** (0.042)	0.467	0.431*** (0.040)	0.446	0.524*** (0.042)	0.530	0.467*** (0.041)	0.469
Being an ELA teacher in Year 3	0.224* (0.106)	0.089	0.521*** (0.120)	0.182	0.233*	0.093	0.754*** (0.112)	0.261
Being a female	0.058 (0.088)	0.023	0.015 (0.097)	0.005	0.026 (0.094)	0.010	-0.126 (0.103)	-0.043
Years of working at the current school up to Year 3	-0.01* (0.005)	-0.070	-0.001 (0.006)	-0.009	-0.003 (0.006)	-0.015	0.004 (0.007)	0.020
Being a coach/teacher consultant in Year 3	0.018 (0.155)	0.004	0.083 (0.172)	0.016	0.106 (0.130)	0.027	0.225 (0.146)	0.050
Having a Master's degree and higher in Year 3	-0.116 (0.087)	-0.049	-0.075 (0.096)	-0.028	0.218* (0.088)	0.092	0.047 (0.098)	0.017
Perceived pressure on improving student performance on state writing assessment in Year 3	0.023 (0.029)	0.028	0.05 (0.032)	0.054	0.013 (0.028)	0.018	0.031 (0.031)	0.036

Notes: Standard errors are reported in parentheses. * p-value ≤ 0.05 ; ** p-value ≤ 0.01 ; *** p-value ≤ 0.001

REFERENCES

REFERENCES

- --- (2008). Improving teacher efficacy and motivation. http://www.newteachercenter.org/tlcsurvey/
- -- (2009). The MetLife survey of the American teachers: Collaborating for students' success, effective teaching and leadership. http://www.metlife.com/assets/cao/contributions/foundation/american-teacher/MetLife_Teacher_Survey_2009.pdf
- Abelson, R., A. Bernstein. 1976. A computer simulation model of com-munity referendum controversies. Public Opinion Quart. 27 93- 122.
- Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization Science*, 8(3), 289-309
- Ball, D. L. (1996). Teacher learning and the mathematics reforms: What we think we know and what we need to learn. *Phi Delta Kappan*, 77(7), 500–508.
- Borko, H., & Putnam, R. (1995). Expanding a teachers' knowledge base: A cognitive psychological perspective on professional development. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 35–66). New York: Teachers College Press.
- Bryk, A. S., & Schneider, B. (Eds.). (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage Foundation.
- Burt, R. S. (1982). *Toward a structural theory of action: Network models of social structure, perception, and action.* New Work, NY: Academic Press.
- Chudgar, A., & Luschei, T. F. (2009). National income, income inequality, and the importance of schools: A hierarchical cross-national comparison. *American Educational Research Journal*, *46*(3), 626-658.
- Cobb, P., McClain, K., Lamberg, T. d. S., & Dean, C. (2003). Situating teaches' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*(6), 13-24.
- Coburn, C. E. (2001). Collective Sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23* (2). 145-170
- Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, 30(3), 203-235.
- Cohen, D. K., & Ball, D. (1990). Policy and practice: An overview. *Educational Evaluation and Policy Analysis*, 12(3), 347–353.

- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, Instruction, and Research. *Educational Evaluation and Policy Analysis*, 25(2), 119-142.
- Cohen, D., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, *102*, 294–343.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin* 112: 155–159.
- Coleman, J.S. (1998). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, (Supplement 1998): S95-S120.
- Cook, T. D., Shadish, S., & Wong, V. A. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from withinstudy comparisons. *Journal of Policy and Management*. 27 (4), 724–750.
- Corcoran, T. (1995). *Helping teachers teach well: Transforming professional development* (CPRE Policy Brief). New Brunswick, NJ: Consortium for Policy Research in Education. Retrieved November 11, 2003, from <u>http://www.cpre.org/Publications/rb16.pdf</u>
- Cordingley, P., Bell, M., Rundell, B., & Evans, D. (2005). *The impact of collaborative continuing professional development (CPD) on classroom teaching and learning*, from http://nationalstrategies.standards.dcsf.gov.uk/files/downloads/pdf/09598003e49523abff7 94962e2752c81.pdf
- Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis*, 29(4), 262-295
- Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *The Phi Delta Kappan*, 76(8), 597-604.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and Abroad: National Staff Development council.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*, 181-199.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81-112.
- Duflo, E., & Kremer, M. (2003). Use of Randomization in the Evaluation of Development Effectiveness1, World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness. Washington, D.C.
- Firestone, W. (1996). Images of teaching and proposals for reform: A comparison of ideas from cognitive and organizational research. *Educational Administration Quarterly*, 32(2), 209– 232.
- Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods and Research, 29*, 147-194.
- Frank, K. A. (2009). Quasi-ties: Directing resources to members of a collective. American Behavioral Scientist. 52 (12). 1613-1645
- Frank, K. A. and Zhao, Y. (2005). Subgroups as a Meso-Level Entity in the Social Organization of Schools. Chapter 10, pages 279-318. Book honoring Charles Bidwell's retirement, edited by Larry Hedges and Barbara Schneider. New York: Sage publications
- Frank, K. A., & Fahrbach, K. (1999). Organization culture as a complex system: Balance and information in models of influence and selection. *Organization Science*, *10*(3), 253-277
- Frank, K. A., Zhao, Y., & Borman, (2004). Social Capital and the Diffusion of Innovations within Organizations: Application to the Implementation of Computer Technology in Schools. Sociology of Education, 77, 148-171.
- Frank, K. A., Zhao, Y., Penuel, W. R., Ellefson, N. C., & Porter, S. (2011). Focus, fiddle and friends: A longitudinal study of characteristics of effective technology professional development. *Sociology of Education*, 84(2), 137-156.Frank, K., A., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A.,& McCrory, R. (2008). Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? *Educational Evaluation and Policy Analysis*, 30(1), 3-30.
- Gallagher, H. A., Penuel, W. R., Murphy, R. F., Bosetti, K. R., Shields, P. M., Toyama, Y., et al. (2008). National Evaluation of Writing Project professional pevelopment: Year 1 report. Menlo Park, CA: SRI International.
- Gallagher, H. A., Penuel, W. R., Murphy, R. F., Bosetti, K. R., Shields, P. M., Toyama, Y., et al. (2009). National Evaluation of Writing Project professional pevelopment: Year 2 report. Menlo Park, CA: SRI International.
- Garet, M. S., Cronen, S. Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., zhu, P., Sztejnberg, L., & Silverberg, M. (2008). The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement. Department of Education: Institute of Education Science
- Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. American Educational Research Journal, 38(4), 915-945.
- Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? Economics of Education Review, 29, 432-439.

- Graham, S., & Perin, D. (2007a). Writing next: Effective strategies to improve writing of adolescents in middle and high schools—A report to the Carnegie Corporation of New York. Washington, DC: Alliance for Excellence in Education.
- Graham, S., & Perin, D. (2007b). What we know, what we still need to know: Teaching adolescents to write. Scientific Studies of Reading, 11(4), 313-335.
- Graham, S., & Perin, D. (2007c). A meta-analysis of writing instruction for adolescent students. Journal of Educational Psychology, 99, 445–476.
- Hansen, M.T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly, 44, 82-111.
- Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a new teaching profession* (pp. 165-180). Washington, DC: Urban Institute.
- Hargreaves, A., & Fullan, M. G. (1992). Understanding teacher development. London: Cassell
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, NY: Houghton Mifflin.
- Heckman, James J. (1978). Dummy Endogenous Variables in a Simultaneous Equations System, *Econometrica*, 47: 153-161.
- Copas, J.B. and Li, H.G. (1997). Inference for Non-Random Samples. *Journal of the Royal Statistical Society, Series B (Methodological)*,59 (1), 55-95.
- Manski, C. (1995). *Identification Problems in the Social Sciences*. Cambridge, Ma: Harvard University Press.
- Robins, J., Rotnisky, A., and Scharfstein, D. 2000. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In Halloran, E. and Berry, D. (Eds.). Title (pp. 1-95).
- Hoxby, C. (2003). School choice and school competition: Evidence from the United States. Swedish Economic Policy Review, 10 11-67.
- Ingersoll, R. M., & Smith, T. M. (2004). Do teacher induction and mentoring matter? *NASSP Bulletin*, 88(638), 28-40.
- Jackson, K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. American Economic Journal: Applied Economics, 1(4), 85-108

- Kennedy, M. (1999). Form and substance in mathematics and science professional development. National Institute for Science Education Brief, 3(2), 1–7.
- Kennedy, M. (2005). Inside teaching: How classroom life undermines reform. Cambridge, MA: Harvard University Press.
- Knott, A. M. (2003). Persistent heterogeneity and sustainable innovation. Strategic Management Journal, 24(8), 687-705.
- Lieberman, A., & Wood, D. R. (Eds.). (2007). Inside the national writing project: Connecting network learning and classroom teaching. NY: Teachers College Press.
- McLaughlin, M. W. (2006). Building school-based teacher learning communities: Professional strategies to improve student achievement. New York, NY: Teachers College Press.Monge, P. R., Cozzens, M. D., & Contractor, N. S. (1992). Communication and motivational predictors of the dynamics of organizational innovation. Organization Science, 3(2), 250-274.

National Writing Project. http://www.nwp.org/cs/public/print/doc/about.csp

- Newmann, F., M., King, M. B., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. American Journal of Education, 108(4), 259-299.
- Ni, Y. (2009). The impact of charter schools on the efficiency of traditional public schools: Evidence from Michigan. Economics of Education Review, 28, 571-584.
- Nilakanta, S., & Scamell, R. W. (1990). The effect of information sources and communication channels on the diffusion of innovation in a data base development environment. Management Science, 36(1), 24-40.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2000). Effects of small classes on academic achievement: The results of the Tennessee class size experiment. American Educational Research Journal, 37, 123-151.
- Penuel, W. R., & Gallagher, L. P. (2009). Comparing three approaches to preparing teachers to teach for deep understanding in Earth science: Short-term impacts on teachers and teaching practice. The Journal of the Learning Sciences, 18(4), 461-508.
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. American Educational Research Journal, 44(4), 921-958.
- Penuel, W. R., Frank, K. A., & Krause, A. (2006). The distribution of resources and expertise and the implementation of schoolwide reform initiatives. In S. A. Barab, K. E. Hay & D. T. Hickey (Eds.), Proceedings of the 7th International Conference of the Learning Sciences (Vol. 1, pp. 522-528). Mahwah, NJ: Erlbaum.

- Raykov, T., & Marcoulides, G. A. (2008). An Introduction to Applied Multivariate Analysis. New York: Taylor & Francis.
- Rogers, E. (1995). Diffusion of innovations. New York: The Free Press.
- Romer, P. (1990). Endogenous technological change. Journal of Political Economy, 98(5), S71-S102.
- Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. American Educational Research Journal, 44(2), 252-297.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). Estimating Causal Effects: Using Experimental and Observational Designs. Washington, D. C.: American Educational Research Association.
- Shadish, W. R., & Steiner, P. M. (2010). A primer on propensity score analysis. Newborn & Infant Nursing Reviews, 19-26.
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. Journal of the American Statistical Association, 103(484), 1334-1344.
- Smith, T. M., Desimone, L. M., & Ueno, K. (2005). "Highly qualified" to do what? The relationship between NCLB teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. Educational Evaluation and Policy Analysis, 27(1). 75-109
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. Educational Research and Evaluation, 2(1), 50–80.
- Stein, M. L., Berends, M., Fuchs, D., McMaster, K., Saenz, L., Yen, L., et al. (2008). Scaling up an early reading program: Relationships among teacher support, fidelity of implementation, and student performance across different sites and years. Educational Evaluation and Policy Analysis, 30(4), 368-388.
- Sun, M., Frank, K.A., Penuel, W. R. & Kim, C.M. (2010). Formal Leadership versus Informal Leadership: How Institutions Penetrate Schools. Present at Annual Meeting of American Education Research Association, Friday, April 30 – Tuesday, May 4, Denver, Colorado
- Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. Strategic Management Journal, 17, 27-43.
- U.S. Department of Education (2003). No Child Left Behind: A toolkit for teachers. Retrieved from http://www2.edtrust.org/NR/rdonlyres/C638111D-04E3-4C0D-9F68-20E7009498A6/0/tellingthetruthteachers.pdf

- U.S. Department of Education (2010). Blueprint for reauthorization of the Elementary and Secondary Education (ESEA). <u>http://www2.ed.gov/policy/elsec/leg/blueprint/index.html</u>
- Valentine, J. C. & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.
- Viadero, D. (2009, September 16). Effective teachers found to improve peers' performance. Education Week, 29 (3)
- Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. Review of Educational Research, 79(2), 702-739.
- Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). Professional development in the United States: Trends and Challenges. Dallas, TX: National Staff Development Council and The Stanford Center for Opportunity Policy in Education.
- Wilson, S. (2009). Teacher Quality: Education Policy White Paper. National Academy of Education
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistence, Regional Educational Laboratory Southwest.
- Youngs, P., Frank, K.A., Thum, Y.M., & Low, M. (in press). The motivation of teachers to produce human capital and conform to their social contexts. In T. Smith, L. Desimone, & A.C. Porter (Eds.), Yearbook of the National Society for the Study of Education: Vol. 110. Organization and effectiveness of high-intensity induction programs for new teachers. Malden, MA: Blackwell Publishing.

CHAPTER 3: HOW EXTERNAL INSTITUTIONS PENETRATE SCHOOLS THROUGH FORMAL AND INFORMAL LEADERSHIP?

Introduction

This paper investigates the role of formal and informal leaders in supporting the diffusion of external reforms within schools that aim to change instructional practices. External demands from federal, state, or local sources contribute to the institutional context of the classroom, both constraining and enabling instructional change (Dacin, 1997; Scott, 1995; Elmore, 2000). But external institutions may not penetrate schools uniformly, as local forces within a school, including administrators and teachers, retain some agency in selecting classroom practices (O'Day, 2002; Ingersoll, 2003) that reflect their unique social contexts (Supovitz & Weinbaum, 2008). Among many forces that drive variability in reform diffusion, leadership can have a powerful influence on how people respond to external pressures to change their practices (Schein, 1992; Moolenaar, Daly, & Sleegers, 2010).

Leadership does not inhere in a single role; rather, it is evident that in the enactment of external reforms leadership is distributed as a form of activity that is carried out by multiple actors within the school across a range of situations (e.g. Riggan & Supovitz, 2008, p. 103; Spillane, Halverson, Diamond, 2004). Some of these actors are formal leaders, who are designated by the school formal structure and include principals, department chairs, and instructional coaches. These leaders have the potential to influence other teachers' behavior or belief by the authority attached to their formal roles. Others are informal leaders who do not have any formal leadership role that confers authority from the organization; rather, they are leaders by virtue of the fact that many other colleagues nominate them as influencing their instructional practices. Although a number of studies have focused on distributed leadership (e.g. Spillane,

2006; Spillane, Hallett, & Diamond, 2003; Spillane, Halverson, & Diamond, 2004; Spillane, & Camburn, 2006; Rowan, 1990; Leithwood, Mascall, & Strauss, 2008), few studies have examined the mechanism of formal and informal leadership on instructional change through a professional networking and influence perspective. To fill the gap in the literature, this paper examines, in the process of implementing external reform, a) how formal and informal leaders influence instructional practices and b) which types of instructional practices are most responsive to which types of leaders.

The context of my study is the implementation of new policies regarding the inducements to schools to adopt research-based reading strategies in the context of accountability-based reforms implemented as part of No Child Left Behind 2001 (NCLB). Leadership within schools may be especially important in adopting instructional strategies as part of this reform, because external mandates for accountability prescribe outcomes but not means for change. Accountability-based reform seeks to tighten the coupling between the formal structure of schools and the technical core of teaching by holding schools accountable for student outcomes (Elmore, 2000; Rowan, 2006; Spillane & Burch, 2006). While the consequences for poor performance are formally prescribed, the ways of designing the internal implementation processes to shape the outcomes of interests — changes in instructional practices — are left up to each school to navigate (Hess & Petrilli, 2006; O'Day, 2002). In this context, intensive instructional leadership related specifically to implementation monitoring and support, as well as strong professional community, can become crucial in promoting reform (Rowan & Miller, 2007).

To probe into the impacts of formal and informal leadership on the change of instructional practices under the background of the implementation of accountability reform, I

analyze longitudinal data on both social interactions and instructional practices of teachers and leaders in nine schools in a single state in the United States. In particular, I use social network analysis to investigate the conjecture that when *NCLB* as a new external source of institutional pressure to change penetrates schools, formal leaders may influence the degree to which teachers adopt general changes to what they teach (i.e., goals for learning) and how they assess learning, while informal leaders may influence specific pedagogical practices. As follows, I will first review literature on how formal and informal leaders' influence on instructional practices and then hypothesize why each type of leaders enact their influence on different instructional tasks based upon theories of teachers' motives.

The Distinctive Influences of Formal and Informal Leaders on Instruction

I will draw on Stein and Nelson's (2003) framework of instructional leadership content knowledge to understand how formal and informal leaders enact influences on which types of instructional practices. Stein and Nelson proposed leadership content knowledge in four layers: The inner two layers include knowledge of teaching and learning of subject matters in the classroom and the outer two layers include knowledge of how to facilitate the teaching and learning. Correspondingly, different levels/types of leaders may exercise separate impacts on instruction given their content knowledge, including pedagogical and subject knowledge, as well as interpersonal skills.

Formal Leaders' Influence

The primary leader with the most formal authority in U.S. schools is the principal who oftentimes shares the instructional leadership role with other formal leaders, such as assistant principals (Weller & Weller, 2001), department chairs (Weller & Weller, 2000; Weller, 2001; Goldberg, 1996; Mayers & Zapeda, 2002), instructional coaches (Coburn and Russell, 2008;

Neufeld & Roper, 2003) or/and teacher mentors (e.g. Youngs, 2002). Formal leaders do not necessarily have substantive subject knowledge or teaching experience, however; the knowledge they deploy to organize teaching and learning in schools is of a different type.

This knowledge corresponds to the two outer layers of instructional leadership knowledge in Stein and Nelson's (2003) framework, which includes knowledge related to strategies of brokering information flow, setting educational goals, and organizing instructional resources. Formal leaders broker information both flowing from outside to inside schools and flowing among teachers within schools (Barnett, 1984, Friedkin & Slater, 1994). They can affect teachers' collaboration and discussion through setting up collective meetings and privileging certain messages over others (Coburn, 2005). Formal leaders also select policy messages to communicate to teachers and establish specific expectations or agenda for teachers' work (Copland, 2003). Additionally, they can network with teachers and community to obtain internal and external support to achieve their goals (Goldring, Crowson, Laird, & Berk, 2003; Copland, 2003; Rusch, 2005). Finally, formal leaders deploy knowledge relevant to allocating budgets to buy instructional materials and coordinating professional development programs to facilitate teachers' instruction and students' learning (Coburn, 2001; Coburn & Russell, 2008).

Informal Leaders' Influence

Beyond formal leaders, there is an emerging focus on informal leadership among regular teachers who do not have any formal leadership roles but who take on leadership tasks. Because these informal leaders are regarded by colleagues excellent teachers with significant teaching expertise who are deserving of respect and who are sought after for advice (York-Barr & Duke, 2004), I speculate that their influences comprise Stein & Nelson's inner two layers of instructional leadership.

Several studies have illustrated that the impact of informal leaders on instruction and student learning had won that of formal leaders (Crowther, Ferguson, & Hann, 2009; Supovitz, 2008; Supovitz, Sirinides, & May, 2010) and can be disseminated via two main ways, including active collaboration on teaching and learning tasks and the development of instructional advice network (Supovitz et al., 2010, p. 36-37). Active collaboration on teaching and learning tasks are identified as the primary means by which teachers affect their peers, particularly through the activities of peer coaching and teacher mentoring in professional development programs, which provide teachers with opportunities to observe each other teaching and examine student work together (Goldstein, 2004; Showers & Joyce, 1996; York-Barr & Duke, 2004). In addition, through instructional advice networks, teachers influence peers when they provide and seek assistance from each other. For example, Frank, Zhao and Borman (2004) described complementary flows of social capital in which novice technology users modified their teaching practices in exchange for support from technology experts. Similarly, Supovitz (2008) studied school reform networks and found that teachers who did not hold formal leadership positions were the primary support of instructional changes and affected the outcomes of school improvement efforts.

Hypotheses Regarding the Motivations of the Distinctive Influences in the Implementation of Accountability Reform

The role of formal leaders has been conceptualized as a *buffer* between external demands and instructional activities within the school (Honig & Hatch, 2004; Rutledge, Harris & Ingle, 2010; Louis, Febey, & Schroeder, 2005). When school is defined as the unit of accountability, formal leaders, as representatives of the school, negotiate with the district central office and other stakeholders about the extent to which external demands fit the school's own teaching goals and

strategies. Most of the formal leaders, such as principals or instructional coaches, are directly hired or/and evaluated by districts or other external agencies (Sun & Youngs, 2009); therefore, formal leaders may be held responsible for implementing the expectations of these external agencies. To demonstrate their conformity to these external demands, formal leaders can mandate the adoption of curriculum materials, of content standards and of student assessment strategies that align to the accountability specifications of what to teach. Formal leader may do so on the belief that these mandates will ensure the eligibility of their schools for the receipt of state and federal funds and the survival of external sanctions, which in turn will demonstrate the effectiveness of their leadership.

Teachers may conform to formal leadership for numerous reasons, the strongest of which is probably because of the sheer coercive power of formal authority. A worker who does not abide by her supervisor's instructions may be fired. This, however, is rare in the case of teachers because of union contracts and tenure policies. But formal leaders have many ways to mobilize informal incentives and sanctions to induce teacher compliance, such as their control for room assignments, teaching schedules and courses, and even parking spaces. Moreover, a teacher may conform to formal leaders as the representative of a community with which she strongly identifies. She might conform to the norms that formal leaders advocate to gain social standing in her school; by conforming, the teacher makes her behaviors legitimate, and thus the teacher is likely to gain more informal support from others in her school (Youngs et al., in press).

Hypothesis 1: Formal leaders influence teachers' general instructional practices associated with facilitating instruction, setting standards, selecting materials, and assessing students.

In contrast to formal leaders, informal leaders may have pedagogical expertise, but because they do not hold formal positions of authority, they do not have to respond directly to outside pressures to maintain legitimacy. Their motives for influencing other peers may stem from the collective sanctions and incentives associated with NCLB accountability system, because to improve their own condition they have to change the pedagogical practices of others that are directly related to their student learning outcomes. For example, a third grade teacher might seek to influence a second grade teacher to emphasize basic reading skills so that all third grade students will have the necessary building blocks for the reading activities the third grade teacher prefers. Moreover, the third grade teacher can build a reputation as a knowledgeable, competent, and helpful colleague. This increases her social capital, a kind of possession that makes her feel trusted by other faculty and fit in the school community (Akerlof & Kranton, 2002; Spillane, Hallett, & Diamond, 2003).

A teacher may conform to the norms of informal leaders to accrue professional effectiveness or/and fit to the local teacher community. Accountability gives the measures for educational outcomes, but not step-by-step guidelines for implementation. Therefore, when it comes to classrooms, teachers still do not know how to achieve these measured outcomes (Cohen, Fuhrman, & Mosher, 2007; Cohen & Hill, 2001). This ambiguous demand increases the uncertainty of the working environment around teachers (Cohen & Moffitt, 2009). This uncertainty makes teachers likely to seek help from or follow the guidance of close colleagues who have expertise in classroom teaching and share the same contexts, regardless of whether that colleague has a formal position of leadership or authority (Kennedy, 2005). Thus a teacher will accept her colleagues' help if she believes colleagues' teaching strategies will contribute to positive outcomes. Or by conforming to close colleagues' norm, a teacher retains her social

standing in the local teacher community which may in turn allow her to continue having informal access to knowledge and support from others in the school.

Hypothesis 2: Informal leaders influence teachers' specific pedagogical practices (e.g., emphasis on teaching basic reading skills), which comprises the inner layers of instruction.

Sample and Measures

Sample

To examine these two hypotheses, I use data from a large-scale, longitudinal project to investigate teachers' implementation of practices associated with NCLB. The original sample includes data collected from a total of 11 elementary and middle schools from eight school districts in California. Nine of these 11 schools were involved in the final data analysis because of missing data, which will be elaborated later in this paper. Administrators and teachers in the selected schools were surveyed four times (2003, 2005, 2007, and 2008). I use data at the last two time points, because they most represent school settings under the NCLB legislation. Moreover, I focus on reading instruction because reading is one of the core subjects targeted by NCLB legislation (Allington, 2006; Miskel & Song, 2004; National Reading Panel, 2000; Schneiberg & Clemens, 2006).

Table 3.1 shows basic characteristics of schools in the sample in 2007-08 school year. The schools included eight elementary and one middle schools, the grade span in each school indicated in the second column of Table 3.1. School size ranged from 288 to 898 with an average of 541. Six schools had a majority non-White student population. The number of full-time equivalent (FTE) teachers ranged from 18.6 to 43 across schools. Four were Title I schools and most of schools in the sample met requirements for Adequate Yearly Progress (AYP) in reading.

Only one sampled school had funded Reading First programs in the district; however, the school itself was not a Reading First school.

Insert Table 3.1 in Appendix 3 about Here

At the fourth wave of data collection the average teaching experience of the sample was up to 13 years, and the mean of years working at the current school was 7.41 (as indicated in Table 3.2). The sampled teachers' relatively longer working experiences in the current schools give this study a great advantage of uncovering teachers' stable relations across years. The majority of the teachers had full certification (advanced professional, regular/standard/probationary) in their main assignment fields.

Insert Table 3.2 in Appendix 3 about here

Measures

Formal and informal leaders

This study aims to identify paths by which formal and informal leaders affect other teachers' instructional activities. I defined a leader as anyone who was listed by another teacher as providing help with reading instruction (a total of 175 teachers were designated as leaders). In the 2008 survey teachers were asked to indicate their nonteaching duties at the school during the 2007-08 school year. I designated 64 formal leaders given their formal roles: five administrators (e.g., principal and assistant principal), two school reform/school improvement coaches or facilitators, 10 reading, literacy, or English program coordinators, 26 master/mentor teachers or

teacher consultants, and 45 committee or team leaders⁷ (Camburn et al, 2003). The other 110 leaders who did not have such formal roles were defined as informal leaders.

As shown in Table 3.3, the average teaching experience of formal leaders was 13.98 years and the mean of years working at the current school was 8.85 years, which were slightly longer than informal leaders who had averaged 12.24 years of teaching experience and averaged 7.22 years of working experience at the current school. One formal leader and four informal leaders did not have full certification in their main assignment fields (advanced professional, regular/standard/ probationary). However, the differences between the formal and informal leaders were not statistically significant.

Insert Table 3.3 in Appendix 3 about here

Dependent variables

General practices of implementing NCLB- related standards, curricula and assessments in 2008. The measure of implementation of NCLB in 2008 was constructed as an index averaging teachers' responses (1="not at all", 2="to a limited extent", and 3="to a great extent") to the question of "Whether NCLB is affecting your work" in the following five areas (α =0.93): "The curriculum materials I use with students," "The curricular activities I use with students," " The content standards to which I teach," "The number of topics I cover in a particular subject area," and "The ways I assess student learning."

Specific pedagogical practices of teaching basic reading skills in 2008. Teaching basic reading skills is one of the key specific teaching practices targeted by NCLB. To measure such

⁷ Some formal leaders have multiple roles.

pedagogical practice, in the 2008 survey, we asked each teacher to rate how often they had students complete a series of activities as part of reading instruction on a five-point scale: 1= "almost never," 2= "1 or 2 times a month," 3= "1 or 2 times a week," 4= "almost every day," and 5= "one or more times a day."⁸ I aggregated nine items into one composite variable (α =0.90), including "Blend sounds to make words or segment the sounds in words," "Read stories or other imaginative texts," "Practice dictation (teacher reads and students write down words) about something the students are interested in," "Use context and pictures to read words," "Clap or sound out syllables of words," "Drill and practice sight words, e.g. as part of a competition," "Use phonics-based or letter-sound relationships to read words in sentences," "Use sentence meaning and structure to read words," and "Practice letter-sound associations."

Independent variables

Prior general practices of implementing NCLB- related standards, curricula and assessments in 2007: Teachers' instructional practices, to some extent, are consistent over time (e.g., Frank et al., 2004). Moreover, the prior practices can be used to approximate the amount of content knowledge or resources a teacher can make available to colleagues. Therefore our measure of the NCLB effect on prior general practices in 2007 is based on the same items and procedures as for the 2008 survey (α =0.92).

Prior specific pedagogical practices of teaching basic reading skills in 2007: To derive the measure of prior specific practices, I measured how often teachers engaged students in activities of learning basic reading skills as part of reading instruction in 2007. The measure included a subset of items from the measure of *implementing basic reading skills in 2008* but

⁸ We considered recoding to days per year, but this exaggerated the most frequent behaviors, skewing the distribution of responses. The original survey scale used here is roughly the log of days per year.

based on the 2007 survey, with slightly different rating scales for each item (1= "not at all", 2= "1 or 2 times", 3= "3 or 4 time", 4= "5 or 6 times", 5= "more than 6 times"). Then I derived a composite variable by taking the mean of items such as "Read stories or other imaginative texts," "Use phonics-based or letter-sound relationships to read words in sentences," "Use context, pictures, and/or sentence meaning and structure to read words," and "Blend sounds to make words or segment the sounds in words" (α =0.87⁹).

Exposure to professional development in 2008. Teachers may change their behaviors based on exposure to external professional development (Cohen & Hill, 2001; Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet et al., 2001). I hypothesized that general practices of NCLB implementation were more likely affected by professional development in areas of "Using achievement data for decision-making," "Strategies for teaching students from different ethnic/ cultural subgroups," "Strategies for teaching English language learners," and "Strategies for teaching students with disabilities," while specific practices of teaching basic reading skills were more likely affected by professional development in reading instruction. I thus developed two measures of the extent to which teachers received professional development, NCLB-related and reading-related. The variables scale from 0 to 3 (0= "None at all", 1= "1-8 hours", 2= "9-16 hours", 3= "more than 16 hours").

Perceived value of NCLB in 2007. Classic innovation diffusion theory suggests that individuals adopt practices based on the perceived value of those practices (Rogers, 1995; Wolfe, 1994). Therefore I controlled for teachers' perceived value of NCLB. Specifically, in our 2007 survey, we asked teachers to rate the importance of the following reform activities for improving

⁹ In the 2008 data, the short version of the measure of focus on basic skills was strongly correlated with the full measure (correlation coefficient $\rho = 0.94$). Therefore this shortened prior measure is sufficient as a measure of prior practices.

student achievement (0= "Not at all important," 1= "Not very important," 2="Neutral," 3= "Somewhat important," 4= "very important"): "Requiring schools to use research-based curriculum materials", " Holding schools accountable for improving achievement of all subgroups at the school", "Giving parents the choice to change schools if the school is failing", and " Giving parents the choice to purchase tutoring services with a school's federal funds if the school is failing". The measure is based on a composite of these items (α =0.70).

Highest grade taught in 2008. Under NCLB, all schools and even Reading First schools preserved a high level of agency for teachers with respect to day-to-day instructional decision-making. Most elementary schools served grades K-5 or K-6, and the program made funding available only for grades K-3, such that teachers of upper elementary level students had more discretion with respect to curriculum and instruction. Therefore I controlled for highest grade taught.

I also included other variables of teachers' background characteristics in our initial data analysis, such as teaching experience, certification status, and others. However, none of them was close to statistical significance; thus, I dropped them from the final models.

Data Analysis

To examine our hypotheses of the ways in which the impact of formal leaders differs from the impact of informal leaders on instructional changes, I estimated one model for general practices of implementing *NCLB* and another for specific pedagogical practices of teaching basic reading skills. The dependent variables were examined as functions of interactions with both formal and informal leaders (Frank &Fahrbach, 1999) after accounting for individuals' prior practices, exposure to professional development in 2008, perceived values of NCLB in 2007, and highest grade taught in 2008.

The key to our models is to approximate teachers' exposure to formal and informal leaders' influence through professional interactions. I followed Frank et al. (2004)'s approach and defined exposure as a function of the extent of interaction between two teachers (approximated by the frequency of interaction), the type of norms conveyed through help (approximated by the prior practices), and the level of interpersonal skills (approximated by the total number of colleagues helped). For example, assume Bob indicate receiving help from three formal leaders: Lisa weekly (3) who had a prior NCLB implementation of 3 and had been nominated by two colleagues as help providers (2), Tom monthly (2) who had a prior NCLB implementation of 3 and had only been nominated by Bob (1), and Alice daily (4) who had a prior NCLB implementation of 1 and had been nominated by three colleagues as help providers (3). Then, Bob's exposure via Lisa is $3 \times 3 \times 2=18$, via Tom is $2 \times 3 \times 1=6$, and via Alice is $4 \times 1 \times 3=12$. Bob's exposure to his formal leaders' norm is (18+6+12)/3 = 12. More formally, exposure to formal leaders is specified as:

Direct exposure formal leaders' influence_i

n.

$$= \frac{1}{J} \sum_{\substack{j=1, \\ j \neq i}}^{n_i} (Help_{ii'}) \times (Providers' prior implementation_j) \times (Total number of others helped_j)$$

(3.1)

Where J = the total number of formal leaders from whom teacher *i* received help; $i \neq j$ means that help from teacher *i* herself or himself is not counted in the summation. Here, I weighted by the total number of others helped to reflect the fact that more popular teachers may more strongly convey norms (cf. Frank et al., 2004). Given the complex metric of the exposure term, I will report results associated with exposure in units of standardized regression coefficients in the next section. The measure for teacher i 's exposure to informal leaders' influence is constructed in the same way and can be specified as:

Direct exposure informal leaders' influence_i

$$=\frac{1}{I'}\sum_{\substack{i'=1,\\i\neq i'}}^{n_i} (Help_{ii'}) \times (Providers' prior implementation_{i'}) \times (Total number of others helped_{i'}) (3.2)$$

Where I' = the total number of informal leaders from whom teacher *i* received help. I flagged the missing values on exposure variables by 0 and included the flags in the analysis (Cohen & Cohen, 1983; Frank et al., 2008).

Because I included two exposure variables (exposure to formal leaders' influence and exposure to informal leaders' influence) in the model, possible multicollinearity between these two effects had to be considered (Doreian, 1989). In order to analyze the main effects of these two predictors properly, I first added each exposure variable separately into the model along with covariates to generate models I and II in both Tables 4 and 5. Next, I added both exposure terms to the model with the covariates, generating model III in both Tables. The standard errors of these exposure predictors did not change significantly from Model I and Model II to the Model III; therefore, I concluded that multicollinearity between the two influence variables was not statistically substantial.¹⁰

Since I analyzed data at two time points, the high turnover of faculty in these 11 schools between 2007 and 2008 led to a large amount of missing data in the analysis, which featured in

¹⁰ Multicollinearity is a problem of highly correlated or interrelated predictors, which leads to the difficulty of determining the relative importance of formal leaders' influence versus informal leaders' influence. If multicollinearity would have existed, the standard errors for the influence terms in model III (in both Table 3.4 and Table 3.5) would have had a very larger standard errors relative to models I and II. However, this is not the case as shown by the data.

the analysis and interpretation of results¹¹. In the final analysis, a total of 137 cases were used to model general practices of implementing NCLB and 149 cases were used to model specific pedagogical practices of teaching basic reading skills in nine schools.

Last but not least, I controlled for teachers' prior practices to 1) account for pre-existing differences in the outcome variables of interest, which would reduce possible bias in drawing any causal inferences (e.g., Cook et al., 2008; Shadish et al, 2008) ; and 2) to improve the precision of estimates of leaders' influence. To further increase the power of the estimation model, I included significant school fixed effects in the modeling.

Results

Estimating Effects on General Practices of Implementing NCLB- Related Standards, Curricula and Assessments

As model-III in Table 3.4 shows, The estimate of exposure to formal leaders' influence is on the boarder of statistical significance level of 0.05 with the unstandardized coefficient of 0.003 and standardized coefficient of 0.172 (t-value=1.91, p-value \leq 0.059). This suggests the possibility that interactions with formal leaders would positively affect teachers' general practices of implementing NCLB-related instructional standards, curricula, and assessments in 2008, which to some extent supports the first hypothesis.

In contrast, informal leaders had a significantly negative influence (unstandardized coefficient=-0.002, standardized coefficient = -0.209, t-value= -2.75, p-value=0.007). Even if I

¹¹ I compared teachers' characteristics between the 2007 sample and the 2008 sample. On average, teachers in the 2008 sample had one year longer working experience than teachers in the 2007 sample. There were no significant differences in the percentage of teachers who had full certification between these two years of sample. Therefore, I can tentatively conclude that teachers in the 2008 sample represent the same group of teacher in 2007 sample in terms of measured individual background characteristics. However, I found that teachers who had partial certification or who had less teaching experience in 2007 were more likely to leave in 2008.

excluded other important predictors in the model, the effect of exposure to informal leaders was still negative. I will propose possible explanations of this negative effect in the discussion section. Own prior implementation of NCLB in 2007 was the strongest predictor with a standardized coefficient of 0.446. None of other covariates, such as exposure to NCLB-related professional development, perceived value of NCLB, or highest grade taught, was statistically significant.

Insert Table 3.4 in Appendix 3 about Here

Estimating Effects on Specific Pedagogical Practices of Teaching Basic Reading Skills

As shown by model-III in Table 3.5, while interacting with formal leaders did not statistically significantly affect the likelihood of teaching basic reading skills in 2008, interacting with informal leaders did. One-standard-deviation increase in exposure to informal leaders' influence would result in 0.156 standard deviation of increase in teaching basic reading skills in 2008. Comparing the standardized coefficients, the effect of exposure to informal leaders' influence is slightly larger than the effect of exposure to reading-related professional development in 2008 (b=0.143) and about one half of the effect of teachers' own prior specific pedagogical practice of teaching basic reading skills (b=0.308).

In addition, exposure to reading-related professional development and perception of high values of the NCLB promote the practices of teaching basic reading skills. Moreover, as I predicted, teachers who taught the lower grades perceived more pressure to teaching basic reading skills than colleagues who taught higher grades.

I then use Frank's calculations to quantify the robustness of the inference of informal leadership on the change in teaching basic reading skills due to any omitted confounding variable (Frank, 2000). The impact of a confounding variable would have to greater than 0.041 to invalidate this inference, and this unmeasured confounding variable would have to be correlated with the outcome variable of teaching basic reading skills at 0.194 and with exposure to informal leaders' influence at 0.293. These are medium correlations in social science according to Cohen's criterion¹² and imply a medium-level robustness of inference. It is intuitive to compare this impact to that of a measured covariate. Partialling for schools as fixed effects and prior status of teaching basic reading skills, one of the strongest covariates is the variable of received reading-related professional development, with an impact of (0.038=0.22×0.17). Thus the impact of an unmeasured confound necessary to invalidate the inference would have to be stronger than the impact of reading-related professional development.

Insert Table 3.5 in Appendix 3 about here

Discussion

This study addresses the need to examine how formal and informal leaders promote instructional changes when external institutions penetrate schools. As informed by theories of both distributed leadership and social influence processes, I modeled how teachers' instructional practices were influenced through interactions with formal and informal leaders. Findings in this study have several theoretical and practical implications, yet limitations.

¹² Cohen's benchmarks are generic descriptors of the magnitude of effect size. Studies in education are likely to smaller effect sizes than other areas (Valentine & Cooper, 2003). Therefore, using Cohen's labels may be misleading and needs to be cautious.

Theoretical Implications

This study contributes to the literature on distributed leadership by providing tentative evidence to support the hypothesis that when the institution of NCLB penetrates schools, formal leaders affect general practices of setting standards, selecting materials, and assessing students. Although I lacked direct measures of leaders' motivation, I also hypothesize the motives of leaders to diffuse external messages based on their content knowledge and local niches. For example, formal leaders may have the skills and resources of facilitating teaching and learning. Their primary responsibility for leading instructional improvement efforts makes them concerned about and motivated to convey expectations on general instructional practices in alignment with NCLB expectations. In contrast, informal leaders positively affect specific pedagogical practices for teaching basic reading skills, hypothetically because they possess subject knowledge of classroom instruction and their primary teaching responsibility in the school system makes them an ideal conduit for conveying normative pressures on classroom instruction.

The ways in which instructional leadership activities are distributed across formal and informal positions unveiled in this study challenge the traditional loose-coupling theory of school organization but also reveal other forms of loose-coupling in the diffusion of an external institution. Traditional loose-coupling theory posited that decisions about what should be taught, how to teach, and how to evaluate teaching and learning, tend to reside in individual classrooms, rather than in the "principal's office" (Bidwell, 1965; Weick, 1976). However, I found that under the penetration of NCLB, school leaders did try to transform the technical core of instruction to align with NCLB expectations. Rather than loosely coupled, the "principal's office" affects what goes into instruction when classroom teachers sought help from formal leaders and adjusted their instructional practices according to formal leaders' influence. However, loose-coupling still

characterizes the degree to which policies can influence adoption of specific instructional strategies. Formal leaders did affect the choice of curriculum materials, establishment of teaching standards and assessment of student learning, but they had a limited influence on how teachers taught basic reading skills in individual classrooms.

I further hypothesize that the non-uniform penetration of NCLB institutions through formal and informal leadership is a mixed result of not only the variation in professional influences experienced by individual school actors but also of the loose-coupling of the structure of accountability policy. The initial wave of reform was built on a strong accountability system that included inducements to adopt "research-based" reading strategies. In some cases, as in the case of Reading First, schools were mandated to adopt these strategies. However, in the schools of our sample, there was no mandate, and any influence of the policy on teaching reading practices would have to be interpreted as normative in nature. When the overarching NCLB legislation approached to schools but not with implementation strategies, individuals within schools had opportunities to choose how to interpret this legislation and develop diverse tasks to implement these norms based on their own interests and school existing practices. The loosecoupling between overarching NCLB norms and implementation strategies developed at local level intensified the differentiation in individual teachers' adoption processes. Since such loosecoupling structure of NCLB legislation, it takes time for such policy to gain legitimacy among teachers. Before all teachers buy in this policy, although some teachers (such as informal leaders) in the school may start to implement it, other teachers in the schools may still resist adopting this external mandate and can take the "free-ride" under the collective incentive and sanction. That might explain the negative impact of informal leaders' influence on general practices of adopting NCLB related standards, curriculum and assessments.

Beyond these theoretical contributions this study adds methodological value to the emerging interest in using social network data and analytical strategies to provide direct evidence of the effects of educational leadership on teaching practice (e.g., Moolenaar et al.,2010; Spillane, Healey, & Kim, 2010; Penuel et al., 2010). Rather than simply describing the characteristics of networks (Moolenaar et al., 2010), I used sophisticated estimation strategies to predict the magnitude of leaders' networking effects on instructional changes using longitudinal data.

Practical Implications

This study highlights the potential for non-uniform diffusion of external reform within schools, which is not a new concept (Cohen & Hill, 2001). Beyond reemphasizing on how the school contexts mediate the successful implementation (e.g. McLaughlin, 1987; Spillane et al., 2000), this study focuses on the role of distributed leadership. If informal leaders' influence on teaching basic reading skills was consistent with formal leaders' influence on general practices of implementing NCLB, teachers would receive coherent guidance on improving reading instruction and further student reading achievement. Otherwise, the non-uniform diffusion processes through formal and informal leadership would exacerbate the loose-coupling structure between external intervention and school practices (Cohen & Ball, 1999).

To lead the successful implementation of external reforms at local school level, a strong team needs to recognize the impacts of both formal and informal leaders. Specific pedagogical activities may directly affect student learning. But standards, materials, and assessment that define the local contexts of teaching may affect student learning as well. Thus, building a strong and collaborative leadership team that includes both formal leaders and teacher leaders is crucial for the successful penetration of an external institution. Oftentimes, it is obvious to identify

formal leaders; but we haven't paid enough attention to these informal leaders who are not supported by formal authority but have intense and productive interactions with other teachers (Frank et al., 2008). To leverage teachers' interaction to facilitate policy implementation, the school principal is recommended to know which teachers inhabit the informal role of providing professional help and how the social structure in the school is formulated.

Beyond recognizing the significance of both formal and informal leadership, several personnel management strategies are suggested to develop effective leaders. First, professional development programs can emphasize different content for different groups of school staff. Formal leaders need to have clear and sufficient information on how to facilitate teaching and learning. However, informal leaders need relatively more support on specific content knowledge and teaching methodologies. Second, different role expectations across the spectrum of school faculty are expected to be clarified in personnel evaluations. We may expect senior teachers with instructional expertise to not only be good at their own teaching, but also help other teachers and lead instructional reform, which can be included in their job description and annual commitment (Frank et al., 2008). Third, in alignment with job expectations, teachers should be compensated and been given incentives for disseminating instructional expertise. For example, compared to an individual-based performance compensation system, group-based performance compensation system would be theoretically expected to be more effective to promote the normative influence and diffusion of knowledge among school staff (Kelley & Protsik, 1997).

Limitations

This study has three key limitations. First, I have analyzed existing social relations in school organizations, which allowed me to describe the stable social structure and to estimate outcomes given on interactions. However, these data did not indicate who initiated the helping

relation. I propose that future studies can explore this question either through collecting empirical data on with whom teachers want to interact or employing simulation techniques such as agent-based modeling (Wilensky & Resnick, 1999 – see Coburn 2005). Second, I only included teachers from one single state in this analysis; therefore, findings from this study have limited generalizability to the population of public schools in the United States. Moreover, the leadership dynamics is examined under the implementation NCLB, which has unique structure and tasks related to accountability. Therefore, some findings may be limited to this context. Third, this study only includes data on adjacent two years. I was not able to examine how school collaborative norm may mediate leaders' influence on instruction through provision of instructional help. Future studies should examine the dynamic changes of collaborative norms among teachers over a longer period and how formal and informal leadership would shape collaborative norms, as well as instruction and learning.

Conclusion

Accountability reform in the implementation process of NCLB legislation is one of the major political efforts in American education that have focused on improving the outcomes of classroom instruction. This external institution of schooling has not only highlighted the school formal leaders' role in promoting instructional changes but also activated other regular teachers' leadership roles (Camburn, Rowan, & Taylor, 2003; Elmore, 2000). Formal leaders facilitate teaching and learning through influencing general instructional practices while informal leaders enact their influence on specific classroom interactions between teachers, students, and materials. Such split but complementary normative influences require policymakers' attention in recognizing the non-uniform local implementation through school leadership. Despite the limitations, this study paves the way for future studies to examine the configuration of

instructional leadership roles and to design personnel management strategies (e.g., professional development, compensation, and evaluation) that develop effective leaders.

APPENDICES

APPENDIX 3

ID	Grade	Student	White	FTE	Title I	Met	Reading
	Span	Enrollmen		Teachers	School	AYP?	First
	-	t					District
Pomo (1)	K-5	441	56.0%	25	No	Yes	No
Pasteur (3)	K-6	898	0.7%	43	Yes	No	No
La Plaza	V C	542	14.6%	27	No	Yes	Yes
Charter (8)	K-0						
Glade (14)	K-8	646	0.3%	29	Yes	No	No
Forest (26)	K-8	538	27.1%	26.8	Yes	Yes	No
Crosswinds	V 5	619	37.6%	33.3	No	Yes	No
(39)	K-3						
Hermosa	5 0	554	70.6%	22.2	No	Yes	No
(48)	5-8						
Sage (53)	K-4	342	64.6%	19.2	No	Yes	No
Dickersen	T Z C	288	25.7%	18.6	Yes	Yes	No
(54)	К-Э						

Table 3.1 School Demographic Information in 2007-08

Notes:

 In this column of AYP status, "Yes" means that the school met AYP in both reading and math in school year of 2007-08; "No" means that the school did not meet AYP in either reading or math except that Pasteur (school #3) did not meet AYP in reading but met AYP in math.
Data sources: Common Core of Data from the National Center for Education Statistics for the 2007-08 school year; Reading First Eligible District from California Department of Education 3. All school names in the table are pseudonyms.

Table 3.2 Teacher Demographics from 2008 Survey

	Characteristics Of Only Nominators	Characteristics of All Faculty (n=228)
Variables	(n=168)	(11 220)
Working experience $(n=168^{a})$		
Mean of years teaching	13	13.09
Mean of years working at the current school	7.47	7.41
Teacher credential status $(n=168^{a})$		
Percentage of partial certification (temporary, provisional, or emergency state certificate)	3 (1.79%)	26 (11.4%)
Percentage of full certification (advanced professional, regular/standard /probationary)	165 (98.21%)	202 (88.6%)

Note: ^a The sample includes all teachers that received help from others and were involved in the final data analysis.

Table 3.3 Demographic Characteristics of Formal and Informal Leaders

	Formal	Informal	Not being
	Leaders	Leaders	nominated
Variables	(n=64)	(n=110)	(n=54)
Working experience			
Mean of years teaching	13.98	12.24	13.65
Mean of years working at the current school	8.85	7.22	6.15
Teacher credential status			
Number and percentage of partial certification	1	4^{a}	1
(temporary, provisional, or emergency state certificate)	(1.56%)	(4.44%)	(1.85%)
Number and percentage of full certification	63	86 ^a	53
(advanced professional, regular/standard /probationary)	(98.44%)	(95.56%)	(98.15)
Expertise as approximated by prior practices			
Mean of prior general practices of implementing NCLB-related standards, curricula and assessments in 2007	1.09	1.26	0.99
Mean of prior specific pedagogical practices of teaching basic reading skills in 2007	3.77	3.57	3.00

Note: ^a 20 cases were missing on this measure. On these measures, there were no statistically significant differences between formal and informal leaders; there were no statistically significant differences between leaders and those who were not nominated by others.

	Model-I		Model-II		Model-III	
	Unstanda rdized coefficie	Standardi zed coefficie	Unstanda rdized coefficie	Standardi zed coefficie	Unstanda rdized coefficie	Standardi zed coefficie
	nt	nt	nt	nt	nt	nt
Prior general practices of implementing NCLB-related standards, curricula, and assessments in 2007	0.489*** (0.08)	0.472	0.451*** (0.079)	0.436	0.461*** (0.079)	0.446
Exposure to formal leaders' influence on implementing NCLB- related standards, curricula, and assessments	0.003 (0.002)	0.146	_	_	0.003 ^a (0.0016)	0.172
Exposure to informal leaders' influence on implementing NCLB related standards, curricula, and assessments	_	-	-0.002* (0.001)	-0.198	-0.002** (0.0008)	-0.209
Exposure to NCLB- related professional	0.045 (0.084)	0.036	0.025 (0.087)	0.020	0.024 (0.086)	0.019

Table 3.4 Estimating General Practices of Implementing NCLB-Related Standards, Curricula, and Assessments in 2008

Note: N=137

2008

development in 2008

Highest grade taught in

Perceived value of

NCLB in 2007

Model-I includes the effect of formal leaders' influence, while model-II includes the effect of informal leaders' influence. Model-III contains partial effects of formal leaders' and informal leaders' influence, after controlling for covariates.

-0.044

0.059

-0.033

(0.053)

0.013

(0.023)

-0.046

(0.053)

0.016

(0.023)

-0.06

0.051

-0.043

0.043

*p-value <0.05, ** p-value <0.01, ***p-value <0.001

-0.034

(0.054)

0.018

(0.024)

^a t-value=1.91, p-value=0.059

	Model-I		Model-II		Model-III	
	Unstandar dized coefficient	Standard ized coefficie nt	Unstandar dized coefficien t	Standard ized coefficie nt	Unstanda rdized coefficien t	Standardi zed coefficie nt
Prior specific pedagogical practices of teaching basic reading skills in 2007	0.27*** (0.057)	0.313	0.262*** (0.056)	0.304	0.267*** (0.056)	0.308
Exposure to formal leaders' influence on teaching basic reading skills	0.002 (0.0005)	0.117	_	_	0.0004 (0.0008)	0.041
Exposure to informal leaders' influence on teaching basic reading skills	_	_	0.002** (0.0005)	0.243	0.001* (0.0004)	0.156
Exposure to reading- related professional development in 2008	0.14* (0.056)	0.145	0.136* (0.055)	0.14	0.139* (0.055)	0.143
Perceived value of NCLB in 2007	0.066 (0.056)	0.068	0.112* (0.053)	0.116	0.157** (0.056)	0.160
Highest grade taught in 2008	-0.123*** (0.029)	-0.309	-0.122*** (0.028)	-0.305	-0.120*** (0.028)	-0.302

Table 3.5 Estimating Specific Pedagogical Practices of Teaching Basic Reading Skills in 2008

Note: N=149

Model-I includes the effect of formal leaders' influence, while model-II includes the effect of informal leaders' influence. Model-III contains partial effects of formal and informal leaders, after controlling for covariates.

*p-value <0.05, ** p-value <0.01, ***p-value <0.001

REFERENCES
REFERENCES

- Akerlof, G. A., & Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4), 1167-1201.
- Allington, R. L. (2006). Reading lessons and federal policy making: An overview and introduction to the special issue. *Elementary School Journal*, 107: 3-15.
- Barnett, B. G. (1984). Subordinate teacher power in school organizations. *Sociology of education*, 57 (January). 43-55
- Camburn, E. Rowan, B. & Taylor. J. E. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis*, 25(4). 347-373
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211-244.
- Coburn, C. E. (2005) Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy*, *19*(3), 476-509.
- Coburn, C. E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation in schools. *American Educational Research Journal*, 43(3), 343-379.
- Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, 30(3), 203-235.
- Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.
- Cohen, D. K., Fuhrman, S. H., & Mosher, F. (2007). Conclusion: A review of policy and research in education. In S. H. Furhrman, D. K. Cohen & F. Mosher (Eds.), *The state of education policy research* (pp. 349-382). New York: Routledge.
- Cohen, D.K., & Moffitt, S. L. (2009). *The Ordeal of Equality: Did Federal Regulation Fix the Schools?* Harvard University Press.
- Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral science.(Second Edition)*. New Jersey: Lawrence Erlbaum Associates, Publishers

- Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology* Supplement 94: S95-S120.
- Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis*, 25(4), 375-395.
- Crowther, F., Ferguson, M., & Hann, L. (2009). *Developing teacher leaders: How teacher leadership enhance school success*. Thousand Oaks, CA: Corwin Press.
- Dacin, T. M. (1997). Isomorphism in context: The power and prescription on institutional norms. *Academy of Management Journal, 40*(1), 46-81.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81-112.
- Doreian, P. (1989). Two Regimes of Network Autocorrelation, pp 280-295. In M. Kochen (Ed.) *The Small World*. Norwood: Ablex
- Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington, D. C.: The Albert Shanker Institute.
- Elmore, R. F. (2002). Bridging the gap between standards and achievement: The imperative for professional development in education. Washington, DC: The Albert Shanker Institute.
- Elmore, R. F. (2005). Accountable Leadership. The Educational Forum, 69 (2), 134-142
- Frank, K. A. & Fahrbach, K. (1999). Organization culture as a complex system: Balance and information in models of influence and selection. *Organization Science*, *10* (3), 253-277
- Frank, K. A., Zhao, Y., & Borman, (2004). Social Capital and the Diffusion of Innovations within Organizations: Application to the Implementation of Computer Technology in Schools. Sociology of Education, 77, 148-171.
- Frank, K., A., Sykes, G., Anagnostopoulos, D., Cannata, M., Chard, L., Krause, A., (2008). Does NBPTS certification affect the number of colleagues a teacher helps with instructional matters? *Educational Evaluation and Policy Analysis*, 30(1), 3-30.
- Friedkin, N. E. & Slater, M. R. (1994). School leadership and performance: A social network approach. *Sociology of Education*, 67 (2), 139-157
- Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.

Goldberg, M. F. (1996). A school without chairs. Phi Delta Kappan, 78(4), 327-329.

- Goldring, E. B. (1990). Elementary school principals as boundary spanners: Their engagement with parents. *Journal of Educational Administration*, 28(1), 53-62.
- Goldstein, J. (2004). Making sense of distributed leadership: The case of peer assistance and review. *Educational Evaluation and Policy Analysis*, 26(2), 173-197.
- Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness & School Improvement*, 9 (2), 157-202.
- Hallinger, P. & Heck, R.H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, *32*(1), 5-44.
- Heck, R. H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal*, *46*(3), 659-689.
- Hess, F. M., & Petrilli, M. J. (2006). *No Child Left Behind*. New York, NY: Peter Lang Publishing, Inc.
- Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, *33*(8), 16-30.
- Ingersoll, R. M. (2003). *Who controls teachers' work?* Cambridge, MA: Harvard University Press.
- Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational Administration Quarterly*, 33(4), 474-505.
- Kennedy, M. M. (2005). *Inside teaching: How classroom life undermines reform*. Cambridge, MA: Harvard University Press.
- Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly*, 30, 498-518.
- Leithwood, K., & Mascall, B. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly*, 44(4), 529-561.
- Leithwood, K., Mascall, B., & Strauss, T. (2008). *Distributed leadership according to the evidence*. New York, NY: Routledge.
- Louis, K. S., Febey, K., & Schroeder, R. (2005). State-mandate accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27(2), 177-204.
- Mayers, R. S., & Zapeda, S. J. (2002). High school department chairs: Role ambiguity and conflict during change. *NASSP Bulletin*, 86(632), 49-64.

- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171-178.
- Miskel, C., & Song, M. (2004). Passing reading first: Prominence and processes in an elite policy network. *Educational Evaluation and Policy Analysis*, 26(2), 89-109.
- Moolenaar, N. M., Daly, A. J., & Sleegers, P. J. C. (2010). Occupying the principal position: Examining relationships between transformational leadership, social network position, and schools' innnovative climate. *Educational Administration Quarterly*, 46(5), 623-670.
- National Reading Panel (2000). *Report of the National Reading Panel: Teaching children to read.* Washington, DC: National Institute of Child Health and Human Development, National Institutes of Health.
- Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacitypromises & Practicalities.* . Washington, D. C.: Aspen Institute Program on Education and the Annenberg Institute for School Reform.
- O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 293-329.
- Printy, S. M. (2008). Leadership for teacher learning: A community of practice perspective. *Educational Administration Quarterly*, 44(2), 187-226.
- Riggan, M., & Supovitz, J. A. (2008). Interpreting, supporting, and resisting change: The geography of leadership in reform settings. In J. Supovitz & E. H. Weinbaum (Eds.), *The implementation gap: Understanding reform in high schools*. NY: Teachers College Press.
- Rogers, E. (1995). Diffusion of innovations. New York: The Free Press.
- Ross, J. A. & Gray, P. (2006). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education*, 29 (3), 798-822
- Rowan, B. (1990). Commitment and control: Alternative strategies for the organizational design of schools. In C. B. Cazden (Ed.), *Review of research in education: Vol. 16* (pp. 353-389). Washington, DC: American Educational Research Association.
- Rowan. B. (2006). The new institutionalism and the study of educational organizations: changing ideas for changing times. In H-D Meyer and B. Rowan (Ed.), *The New Institutionalism in Education*. Albany: State University of New York Press.
- Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal*, 44(2), 252-297.
- Rutledge, S. A., Harris, D. N., & Ingle, W. K. (2010). How principals "bridge and buffer" the new demands of teacher quality and accountability: A mixed-methods analysis of teacher hiring *American Journal of Education*, *116*(2), 211-242.

- Schein, E. H. (1992). *Organizational culture and leadership*. San Francisco, CA: Jossey-Bass Publishers.
- Schneiberg, M., & E. S. Clemens (2006). The typical tools for the job: Research strategies in institutional analysis. *Sociological Theory*, 24: 195-227.
- Scott, R. (1995). Institutions and organizations. Thousand Oaks, CA: Sage.
- Showers, J. & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership*, 53(6), 12–16.
- Spillane, J. (2006). Distributed leadership. San Francisco, CA: Jossey-Bass.
- Spillane, J. P. & Burch, P. (2006). The institutional environment and instructional practice: Changing patterns of guidance and control in public education. In H-D Meyer and B. Rowan (Ed.), *The New Institutionalism in Education*. Albany: State University of New York Press.
- Spillane, J. P. & Camburn (2006). The practice of leading and managing: The distribution of responsibility for leadership and management in the schoolhouse. Paper presented at the Annul Meeting of the American Educational Research Association, San Francisco, April, 2006.
- Spillane, J. P. & Hunt, B. R. (2010). Days of their lives: a mixed-methods, descriptive analysis of the men and women at work in the principal's office. Journal of Curriculum Studies
- Spillane, J. P., Hallett, T., & Diamond, J. B. (2003). Forms of capital and the construction of leadership: Instructional leadership in urban elementary schools. *Sociology of Education*, 76(1), 1-17.
- Spillane, J. P., Healey, K. & Kim, C. (2010) Leading and Managing Instruction: Using social network theory and methods to explore formal and informal aspects of the elementary school organization. In. A. J. Daly (Ed.), *Social Network Theory and Educational Change*, Harvard Education Press:MA, 129-158.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (1999). *Distributed Leadership: Toward a Theory of School Leadership Practice*. Annual Meeting of the American Educational Research Association, Montreal.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Towards a theory of leadership practice: a distributed perspective. *Journal of Curriculum Studies*, *36*(1), 3-34.
- Spillane, J. P. Reiser, B. J. Reimer, T. (2002). Policy implementation and Cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72 (3), Standards-Based Reforms and Accountability, 387-431
- Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. Educational Evaluation and Policy Analysis, 25(4), 423-448.

- Supovitz, J.A. (2008). Instructional leadership in American high schools. In M. M. Mangin & S. R. Stoelinga (Eds.), *Effective Teacher Leadership: Using research to inform and reform*. New York: Teachers College Press.
- Supovitz, J., Sirinides, P., & May, H. (2010). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, 46(1), 31-56.
- Supovitz, J., & Weinbaum, E. H. (2008). Reform Implementation Revisited. In J. Supovitz & E. H. Weinbaum (Eds.), *The implementation gap: Understanding reform in high schools*. New York: Teachers College Press.
- Weick, K. E. (1976). Educational organizations as loosely-coupled systems. *Administrative Science Quarterly* 21(1), 1-19.
- Weller, L. D. (2001). Department heads: The most underutilized leadership position. *NASSP Bulletin*, 85(625), 73-81.
- Weller, L. D., & Weller, S. J. (2000). *Quality human resources leadership: A principal's handbook*. Lanham, MD: Scarecrow Press.
- Weller, L. D., & Weller, S. J. (2001). *The assistant principal: Essentials for effective school leadership.* Throusand Oaks, CA: Corwin Press, Inc.
- Wilensky, U., & Resnick, M. (1999). Thinking in Levels: A Dynamic Systems Approach to Making Sense of the World. *Journal of Science Education and Technology*, vol. 8, no. 1, pp. 3-19.
- Youngs, P. A (2002). *State and District Policy Related to Mentoring and New Teacher Induction in Connecticut*. New York: National Commission on Teaching and America's Future.
- York-Barr, J., & Duke, K. (2004). What do I know about teacher leadership? Findings from two decades of scholarship. *Review of educational research*, 74(3), 255-316.

CHAPTER 4: THE USE OF MULTILEVEL ITEM RESPONSE THEORY MODELING TO ESTIMATE PROFESSIONAL INTERACTIONS AMONG TEACHERS

Introduction

This paper investigates the potential of using multilevel item response theory to estimate the depth of teacher interactions, defined as the propensity of endorsing collaborative relationship with regard to mathematics instruction. Multilevel item response theory (IRT) modeling that integrate traditional item response theory (IRT) into multilevel modeling is not new; rather, it has been developed and explored by researchers who have named it as "multilevel item response theory" (Adams, Wilson, & Wu, 1997; Kamata, 2001; Fox, 2007), or "hierarchical measurement model" (Maier, 2001; 2002), or "random item IRT models" (De Boeck, 2008), or "explanatory item response models" (Briggs, 2008; De Boeck & Wilson, 2004). It has been widely applied to estimate students' skill-based abilities based on their responses on standardized tests. However, none of the prior studies have examined the potential for applying multilevel IRT modeling to estimate teachers' professional interactions and diagnose the quality of the instrument used to collect network data.

Quantitative data on teacher interactions are often collected using social network surveys. For example, in a regular teacher network survey, researchers might ask respondents to nominate who helped them teach mathematics and with which types of instructional tasks ("1"=Yes; "0"=No), including doing mathematics problems together, discussing students' work, sharing instructional materials, and so on. These different types of instructional tasks can be treated as items of a small test and the endorsement of the collaboration on a task between two teachers can be treated as an item response. The network survey data have been used in a wide range of empirical studies, including the investigation of professional learning communities in schools

(Coburn & Russell, 2008; Gallagher et al., 2010), the evaluation of educational reform (Cross et al., 2009; Supovitz, 2009) and the study of innovation diffusion within schools (Frank et al., 2004). The growing popularity of using network studies to inform educational policy and practices demands the development of a psychometrically sound measure of teacher interaction. This study will contribute to this methodological development and has three specific purposes:

- 1. Construct a composite measure of teacher interaction from a set of item responses;
- 2. Provide useful diagnostic information to assess the quality of social network survey instruments;
- Demonstrate the possibility to incorporate predictors in the measurement model to investigate multilevel research questions.

Before applying models to empirical data, I will first introduce network theories that inform the unique data structure and measurement problems that this paper is intended to address.

Conceptual Framework

This study focuses on egocentric network data that are collected to inform the social connections of individuals (egos). Egos are assumed to be independent from each other (as displayed in Figure 4.1, ego A is independent from ego B) and researchers are most interested in the relations between the ego and her/his alters.

Insert Figure 4.1 in Appendix 4.B about Here

The depth of interaction between an ego and an alter varies due to many factors within the nested structure of the egocentric network. As illustrated in Figure 2, responses on multiple items that indicate different collaborative activities (e.g., Item1, Item2, and Item3) together capture the characteristics of a *tie*, the relation between the ego and the alter. Ties are indicated by solid lines in Figure 4.2. The teacher who responded to the survey and nominated other colleagues who helped her with teaching mathematics is called the *ego*, while her colleagues who were nominated by her as help providers are *alters*. Items are nested within ties and ties nested within egos (Wellman & Frank, 2000). Therefore, there are at least three-levels of factors that might affect the depth of teacher interaction: item, tie, and ego. The distinction in these three levels is central to our understanding of the measurement problems.

Insert Figure 4.2 in Appendix 4.B about Here

Differences in item characteristics: Items reflect different types of teacher interactions that may provide teachers with learning opportunities. Given the potential of providing learning opportunities, Coburn and Russell defined teacher interaction in three categories:

Low depth of interactions included talk related to one or more of the following: how to use materials; how to coordinate the text, standards, assessments, and pacing guides, how to organize the classroom; sharing materials or activities; general discussions of how a lesson went or whether students were getting it.

Medium depth of interactions included talk related to one or more of the following: discussions of how lessons went, including a discussion of why; detailed planning for lessons, including a discussion of why; specific and detailed discussion of whether students were learning (but not how students learn); discussion of instructional strategies in the context of observations; doing mathematics problems together with discussion. High depth of interactions included talk related to one or more of the following: pedagogical principles underlying instructional approaches; how students learn, or the nature of students' mathematical thinking; mathematical principles or concepts. (see Coburn & Russell, 2008, p.230)

Teachers involved in low-depth of interactions may simply exchange information or materials. Collaboration in the medium and high depth of interactions requires more than sharing information, but also learning new knowledge about the curriculum, pedagogical strategies, and principles of learning. Such difference in collaborative activities presented by item characteristics should be taken into account in the estimation process.

Differences in tie characteristics: The latent trait of teacher interaction is allocated at this level, which varies randomly given item characteristics. Multiple factors may predict or explain such variation. Frank et al. (2005, 2010) have used homophily (the effect of common attributes on the occurrence of a network relation) and heterophily (interacting with others of different attributes) to investigate with whom and on what tasks teachers collaborate. School actors who have similar characteristics or occupy similar roles may be more aware of each others' needs and strengths and may feel more inclined to help one another (Burt 1982; Frank & Yasumoto, 1998). For example, teachers who teach the same grade or same subjects may share the same group of students, which makes them more likely to interact with each other. Or teachers might interact intensively with others of the same gender or race/ethnicity. In contrast, the heterophily might occur because teachers seek new information or resources by interacting with others of higher status or levels of performance.

Differences in ego characteristics: Individual latent trait of collaboration with others is relatively stable over a short period of time (such as a year), even though expression of specific

types of interactions might reflect local variation in opportunities for successful endorsement of a relation. High levels of the trait increase the probability of virtually every type of collaboration with other members in the school. But this relatively stable trait can also be changed by external interventions, such as receiving high-quality professional development programs that focus on developing teachers' collaborative skills (see Chapter 2 the first sub-study). Or new coaching routines designed by districts can promote teacher collaboration (Coburn & Russell, 2008). Or school principals may facilitate certain interactions (Coburn, 2001). Or state's group incentive strategies invoke more collaboration among teachers than individual performance incentive strategies (Frank et al., 2010; Kelley & Protsik, 1997).

Multilevel Item Response Theory Models

Network theories inform us that to capture the nature of teacher collaboration, we need to develop a measurement model that can account for the differences in item, tie, and ego characteristics and can easily be extended to estimate the effects of temporary/contextual predictors. Network theories also inform us of several limitations of conventional approaches. For example, researchers may simply use the mean or sum of these raw item scores to represent the depth of interactions. In this case, researchers take a risk by assuming that all items are similar in nature and they function equally for all participants. Item characteristics do not enter into the theory of scoring. Moreover, the composite measure often has a very skewed distribution, which limits further statistical analysis (Raudenbush, Johnson, & Sampson, 2003; see also Osgood, Mcmorris, and Potenza, 2002). Traditional IRT takes into account for item characteristics and creates a meaningful metric that appropriately reflects the varying depth of teacher interaction while reducing the skewness of distribution. However, it cannot accommodate the nested structure of network data.

When developing IRT into a multilevel structure, the multilevel IRT (Kamata, 2001; 2002; Fox, 2007) can thus solve several problems that conventional IRT or mean/sum cannot address but are very important for the type of measurement problem in this study. *First*, multilevel IRT can estimate latent traits at different levels simultaneously (Maier, 2001). In this study for example, latent traits may include a) the depth of teacher interaction between a pair of teachers at the tie level and b) the extent to which a teacher is embedded in the network at the individual level. Embeddedness indicates the extent to which teachers are integrated in a dense cluster or multiplex relations of a social network. The more embedded a teacher is, the more resources and constraints of this social network she may face. Being able to estimate these two latent traits is beneficial in educational studies when both network-level contexts and individual-level attributes can help us understand the extent to which teachers can learn best to teach from peers (Cobb, McClain, Lamberg, & Dean, 2003)

Second, multilevel IRT can accommodate dependencies in the nested structure. Interaction is a complicated social action structured by social dependencies. Specifically, within ties, multiple observations on the same tie may be correlated. If teacher *A* helps teacher *B* in preparing curriculum materials, teacher *A* may be also likely to discuss teaching philosophies and the nature of student learning with teacher *B*. Similarly, across ties within individuals, ties initiated by the same individual may be similar. If teacher *A* has the tendency for receiving help, she may be more likely than others to get it from B, C, and all others in her school. The multilevel IRT model accounts for dependencies and allows researchers to assume the conditional independence within the higher-level units (Raudenbush & Bryk, 2002). The model assumes that within ties, multiple observations on interactions between two teachers are

correlated, but conditionally independent across ties. Similarly, within individuals, ties are dependent but conditionally independent across units of individuals.

Third, multilevel IRT can proportion the total variance into different levels, which enables researchers to more accurately estimate standard errors of measurement associated with latent traits (Raudenbush & Bryk, 2002) and the relations between latent traits and predictors (Kamata, 2001; Pastor, 2003). The model proportions the total variance and covariance into separate components at the item, tie and ego levels, which allows one to understand the extent of measurement errors given each level of latent traits. Furthermore, one can develop reliability or an information function to indicate the precision of the test given the latent trait and item parameters.

Fourth, the measurement model can be combined with a structural model by including predictors and covariates at any level when researchers are interested in understanding the relationship between particular factors and the latent trait. Furthermore, by incorporating measurement errors in the model (Maier, 2001) and the partitioning of variance-covariance components into different levels, the multilevel IRT model allows for better estimations of the true relationship between predictors and latent traits (Pastor, 2003; Maier, 2001; 2002). By including person-item covariates, one can also examine if item parameters change as a function of group effects. This would be an example of a model used to analyze Differential Item Functioning (DIF, cf., Briggs, 2008)

Instrument and Sample

This study draws on data from a larger study of Vanderbilt's Middle School Mathematics and the Institutional Setting of Teaching (MIST). On a regular on-line survey, researchers included a question of asking teachers to nominate the names of colleagues from whom they

sought advices and help with regard to 12 different types of instructional matters, as listed in Table 4.1. The presence or absence of interaction between two teachers on a particular facet of instruction was defined as binary, 1 = "Yes" and 0 = "No".

Insert Table 4.1 in Appendix 4.C about Here

Longitudinal data were collected from most of the school faculty members at three time points: 2008-09, 2009-10 and 2010-11. In this paper, I use the 2008-09 network data as an illustration, including 223 middle school mathematics teachers who reported 586 ties on 12 items. Among these teachers, as shown in Table 4.2, about 61.735% were White, 28.718% were black, and less than 10% were other race and ethnicities, such as Asian, Hispanic, Latino, Native American, and others. About 68% were female and 89% of teachers who held full certification (including advanced professional, regular/standard, probationary). On average, teachers had worked in current school for five years and had nine years of teaching experiences in mathematics.

Insert Table 4.2 in Appendix 4.C about Here

Models

Measurement Model

I used Kamata' (2001; 2007; Kamata et al., 2008) unconditional multilevel Rasch model to fit the data to simultaneously estimate item characteristics and the depth of teacher interactions. Level-1 model, item level:

$$\eta_{ijp} = P(Y_{ijp} = 1 | \theta_{jp}, b_i)) = \frac{\exp(\theta_{jp} - b_i)}{1 + \exp(\theta_{jp} - b_i)} = \frac{1}{1 + \exp[-(\theta_{jp} - b_i)]}$$
(4.1)

Where θ_{jp} represents the propensity of endorsing tie *j* within ego *p*'s network; and b_i represents an item difficulty parameter, the probability of endorsing item *i*. I relabel it as item rareness, because the probability of endorsing item *i* might not be totally due to the fact that the task was difficult for teachers to collaborate, but that this task just happened to rarely occur in practice. Y_{ijp} represents the presence or absence of tie *j* nominated by individual *p* on item *i*. Kamata (2001; 2007) showed the transformation of equation (4.1), the probability function, to equation (4.2), the log-linear function. The log-odds of the probability of endorsing item *i* for tie *j* of ego *p*, η_{iip} is modeled as:

$$\log(\eta_{ijp}) = \pi_{0jp} + \sum_{q=1}^{Q-1} \pi_{qjp} X_{qjp}$$
(4.2)

Where in equation (4.2) X_{qjp} is the *q*th item indicator, a dummy variable with a value of 1 when *i*=*q*, and 0 otherwise. Q indicates the total number of items on a test or survey. π_{qjp} is the coefficient associated with X_{qjp} , where q=1, ..., Q-1, the rareness parameter estimate for *i*=*q* equals to $(-\pi_{qjp} - \pi_{0,jp})$ (Kamata, 2001). The Qth item has been coded as the reference item and its rareness is set to be $(-\pi_{0,jp})$.

For the level-2 model of ties, the level-one intercept was allowed to vary across ties with random effects and the difficulty parameters were fixed across ties, which could be simplified as: The Level-2 model, tie level:

$$\pi_{0jp} = \beta_{00p} + u_{0jp}$$

$$\pi_{qjp} = \beta_{q0p} \tag{4.3}$$

Where in equation (4.3) β_{00p} is the intercept for π_{0jp} , while u_{0jp} is a random component of π_{0jp} and assumed to be distributed as N (0, τ_{π}). The π_{qjp} were fixed at level-2 as β_{q0p} , which means that each item has the same rareness estimates across ties. I then modeled β_{00p} as a random effect at ego level.

The level-3 model, ego level:

$$\beta_{00p} = \gamma_{000} + r_{00p} \beta_{q0p} = \gamma_{q00}$$
(4.4)

Where γ_{000} is the intercept of β_{00p} at level-3. $r_{00p} \sim N(0, \tau_{\beta})$ are the random effects of

 β_{00p} . I used the sum of empirical Bayes estimates $(u_{0jp} + r_{00p})$ to represent the latent trait of tie *j* nominated by person *p*, the depth of interaction. β_{q0p} was fixed at level-3 as γ_{q00} .

I grand-mean centered all dummy variables at level-1 to make the coefficient γ_{000} represent the estimate for the reference item. I also used the Laplace approximation algorithm to estimate the model. This approach produces a remarkably accurate approximation to maximum likelihood (ML) and therefore provides efficient (or nearly efficient) estimates of all parameters (see Yang, 1998; Raudenbush, Yang, &Yosef, 2000).

Since this multilevel IRT model is developed from a Rasch model, it also inherits two key assumptions of the traditional Rasch model: unidimensionality and equal discrimination. Unidimensionality: one single parameter of θ_{jp} adequately describes the relation between item rareness and the characteristics of the tie (Reckase, 2009), as indicated in equation (4.1). In other words, we treat θ_{jp} as the common factor of all items (Lord, 1980), a property of the items.

Equal Discrimination: one single rareness parameter describes the characteristics of the items. All items have the same discriminating power, the degree to which item response varies within the level of depth of interaction. But if one assumes that one parameter is not adequate to describe the connection between observed response and the level of interactions, one can use an alternative form of item response function in equation (4.5), where a_i represents item *i*'s discrimination. Compared equation (4.5) to equation (4.1), the Rasch model would be equivalent to setting all a_i to 1.

$$\eta_{ijp} = P(Y_{ijp} = 1 | \theta_{jp}, a_i, b_i)) = \frac{\exp[a_i(\theta_{jp} - b_i)]}{1 + \exp[a_i(\theta_{jp} - b_i)]} = \frac{1}{1 + \exp[-a_i(\theta_{jp} - b_i)]}$$
(4.5)

I tested the extent to which our data met these two assumptions before I fitted the data to the multilevel Rasch Model.

Prediction Model

Once the unconditional model was fit to the data and given that there was significant variation within and between egos (variance component at level-2=1.243, p-value<0.001; variance component at level-3 =2.031, p-value<0.001). To determine if the variation across ties and persons was associated with other predictors, a model including the effects of three predictors was fit to the data. At the tie level, I suspected that a white teacher might feel more comfortable to seek advice from another white teacher, while a non-white teacher might feel more comfortable to seek advice from another non-white teacher. I thus included a variable indicating whether two teachers who had a relation were both white or both non-white at level-2

to examine this race/ethnicity effect. The level-1 model remained the same as equation (4.1) and (4.2), while the second level model was specified as follows:

$$\pi_{0jp} = \beta_{00p} + \beta_{01p} same \ race_{jp} + u_{0jp}$$

$$\pi_{qjp} = \beta_{q0p} \qquad (4.6)$$

Where β_{01p} represents the race/ethnicity effect.

I also wanted to examine whether female versus male teachers or whether white versus non-white teachers had a higher propensity of engaging in deeper collaboration. Therefore I added two dummy variables at the ego level to examine the effects of personal characteristics. The third -level model was then modified as:

$$\beta_{00p} = \gamma_{000} + \gamma_{001} Female_p + \gamma_{002} White_p + r_{00p}$$

$$\beta_{q0p} = \gamma_{q00}$$
(4.7)

Where γ_{001} represents the gender effect after controlling for other covariates in the model. γ_{002} represents the effect of ego's race after controlling for other covariates in the model.

To detect DIF, I added ego-item interactions, "computed as the product of a person predictor (representing a person group) and an item indicator (representing an item or item property)" (Meulders & Xie, 2004, p. 215). The level-1 remained as equation (4.2) and level-2 model remained as equation (4.3), but level-3 was modified as equation (4.8)

$$\beta_{00p} = \gamma_{000} + \gamma_{001} \text{White}_{p} + r_{00p}$$

$$\beta_{q0p} = \gamma_{100} + \gamma_{q01} \text{White}_{p}$$
(4.8)

Where β_{q0p} was predicted by group indicator, such as White *p*, whose coefficient was indicated by γ_{001} and γ_{q01} . The mixed model that combines equation (4.2), equation (4.3) and

equation (4.8) shows clearly how the person-item interactions were included as cross-level interaction effects.

$$\eta_{ijp} = \gamma_{000} + \gamma_{q00} \times White_p \times X_{qjp} + r_{0jp} + u_{00p} \qquad (4.9)$$

I modeled the DIF effect of gender by using the similar method as including $Female_p$ as a predictor at the ego level.

Results

I first investigated how the Rasch model may be applied to the estimation of teachers' professional interactions, testing key model assumptions. Next, I showed the estimates of item rareness parameters and the information functions across estimates of the depth of interactions. I then extended the measurement model to explanatory multilevel models by including DIF indicators and explanatory variables.

Testing Rasch Model Assumptions

Single Dimension

One of the key assumptions of Rasch model is the single dimensionality. One way to satisfy this assumption is to exclude items that do not fit the main dimension. I began tentatively with a two-dimensional notion of the depth of interaction to explore the dimensionality of the observed data based on prior literature (Garrison et al., 2011). The first dimension includes item-1, item-2, item-3, item-4, item-5, and item-7, which describe the underlying construct of discussing underlying principles of teaching and learning. The second dimension includes item-6, item-8, item-9, item10, item-11, and item-12, which describe the other underlying construct of sharing instructional materials and information. To test the null hypothesis that two factors are sufficient, I performed a confirmatory factor analysis (CFA) based on tetrachoric correlation matrix (as shown in Table 4.3). The null hypothesis is the two-factor model fits the observed data

and the proposed model is a plausible description of teachers' responses to these 12 items. This null hypothesis is rejected given several model goodness-of- fit indices. For example, Chi-square statistics of 1930.54 with degree of freedom of 53 is statistically significant at 0.001 level. Standardized Root Mean Square Residual (SRMR) equals to 0.094, larger than the cut-off value <0.08 suggested by Hu and Bentler (1999). Bentler's comparative fit index (CFI) equals to 0.676, which is less than the suggested cut-off value of 0.95 or larger. The estimate of RMSEA is about 0.259, which does not meet the cut-off of < 0.06.

When examining the error variance and R-square¹³ of each item (as indicated in Table 4.4), I noticed that item-8 and item -11 had relatively larger error variances and smaller R-square $(\sigma^2(s.e._8)=0.8 \text{ and } R^2(\text{item-8})=0.2; \sigma^2(s.e._{11})=0.742 \text{ and } R^2(\text{item-11})=0.258)$. The rest of the items are similar to each other. Thus, I removed these two items and performed a one- factor analysis. All goodness-of-fit indices significantly improved in this one-factor model. Using the Likelihood Ratio (LR) test (also shown as chi-square difference test) to compare the one factor model with the two factor model, the LRT statistic is 1930.5397-1296.92=633.616, which is statistically significant when judged against the pertinent chi-square distribution with degrees of freedom of 17 (=53-35, the .001 cutoff for this distribution is 40.79). That is, the one-factor model is significantly better than the two-factor model at the alpha level of 0.001. SRMR changes to 0.086 and CFI changes to 0.74. More comparisons between these two models are included in Table 4.5.

Insert Table 4.3 in Appendix 4.C about Here

¹³ R-square represents the proportion of the variances in the observed items that is accounted for by its corresponding latent variable. It is used as an indicator of each item's common factor reliability (Doll, Raghunathan, & Gupta, 1995)

Insert Table 4.4 in Appendix 4.C about Here

Insert Table 4.5 in Appendix 4.C about Here

I then used Lord's procedure to further confirm that the data after removing item-8 and item-11 contain one dimension (1980, p.21, see also Christofersson, 1975; Muth én, 1977). Lord examined the dimensionality by computing the eigenvalues of the tetrachoric item intercorrelation matrix with estimated communalities placed in the diagonal. If 1) the first eigenvalue is large compared to the second and 2) the second enigenvalue is not much larger than any of the others, then the items are approximately unidimensional. As shown in Figure 4.3, The first eigenvalue equals 6.449 and explaines about 64.49% of total variance. The second eigenvalue equals 0.978, which explains about 9.7% of variance. The third eigenvalue is very close to the second value and equals 0.709, and so forth. As indicated in Table 4.6, the factor loadings on each item are similar to one another. One-dimension after removing item-8 and item-11 seems sensible.

Insert Figure 4.3 in Appendix 4.B about Here

Insert Table 4.6 in Appendix 4.C about Here

Equal Discrimination

I used the procedure recommended by Raudenbush, Johnson, and Sampson (2003) to examine the assumption of equal discrimination by comparing results based on one-parameter and two-parameter models. I used BILOG software (Zimowski, Muraki, Mislevy & Bock, 2003) to estimate a two-parameter model (equation 4.5) and then compared the results with these based on Rasch model (equation 4.1).

In the two-parameter model, item-8 and item-11 had relatively small biserial correlation $(\rho_8=0.418; \rho_{11}=0.485)$, smaller estimates of slopes (a₈=0.892; a₁₁=1.01), and larger standard errors of difficulty estimates (s.e., item8=0.245; s.e, item1=0.147). These two items behaved in a way that appeared different from the other items. I therefore excluded these two items and re-ran the analysis by using the Rasch model. The magnitudes of the biserial correlation coefficients shown in the fifth column in Table 4.7 are close to each other and the standard errors of difficulty estimates are almost equal to each other. The LRT statistics indicate that the Rasch (-2loglikehood = 5993.944) is significantly better than the two-parameter model (-2loglikehood=7153.427). The graphic comparison of item response curves (ICCs) leads to the same conclusion. For the two-parameter model (Figure 4.4 a), the ICCs are similar in shape except item-8 and item-11. In the one-parameter model after removing item-8 and item-11, all ICCs are nearly in the same shape (Figure 4.4 b). The equal discrimination assumption is sensible after removing item-8 and item-11. Therefore, item location parameters can reasonably be interpreted as item rareness, and items and the depth of interactions arguably are calibrated on a common scale.

Insert Table 4.7 in Appendix 4.C about Here

Insert Figure 4.4 in Appendix 4.B about Here

Item Information or Goodness-of- Fit Indices under Multilevel Framework

Item information can be used as an extension of item reliability to describe how well, or precisely, an item measures each level of the latent trait that is being measured by a given instrument. Lord showed (1980) that as a characteristic of maximum likelihood estimators, the reciprocal of the item information was equal to the asymptotic sampling variance of the parameter estimator if the estimator was statistically unbiased (also refer to Reckase, 2009). Because of the inverse relationship between item information and variance, the more information the item contains, the less variance of the item; and vice versa.

Based on Fisher's concept of information, the variance of a maximum likelihood estimator is the reciprocal of the negative expectation of the second derivative of the loglikelihood function with respect to the parameter (Baker & Kim, 2004). Since I used Laplace approximation algorithm to accurately approximate maximum likelihood (ML) estimation in this generalized linear model, I can use the variance component of item parameter generated by HLM software 6.0 as a good proxy of the Fisher's concept of the variance of item parameter and use the reciprocal of the variance component as a good proxy of the Fisher's item information. The x^2 test for item variance can be used as a goodness-of-fit test between the parameter estimate and the observed proportion of endorsement on this item. As indicated in Table 4.8, compared to other items, Item-8 and Item-11 have significantly larger variances than the rest of the items $(\tau_{\pi(8)}=0.014, \text{ p-value of } x^2 \text{ test}=0.004; \tau_{\beta(8)}=14.011, \text{ p-value of } x^2 \text{ test} \le 0.0005; \tau_{\pi(11)}=0.006,$ p-value of $x^2 \text{ test} > .500; \tau_{\beta(11)} = 1.467, \text{ p-value of } x^2 \text{ test} \le 0.0005)$, which indicate less information contained by these two items. The result leads to the same conclusion as elaborated in the previous testing results of single dimensionality and equal discrimination that item-8 and item-11 do not fit the test.

Insert Table 4.8 in Appendix 4.C about Here

Item Rareness

Table 4.9 provides the item rareness estimates, ranked from the most common (1) to the rarest activity (10). Item 4, item-5, item-6, and item-7 are relatively rare collaborative tasks in practice. In contrast, item-2, item-3, item-9, and item-12 are among the most common collaborative tasks.

Insert Table 4.9 in Appendix 4.C about Here

The Distribution of Propensity of the Interactions

Since the score on the level of collaboration given each tie is estimated by the logistic model, the scores estimated as the sum of level-2 residuals and level-3 residuals are in log odds, arranging from -4 to 4. To facilitate interpretation and eliminate negative values of estimates, I added 10 to all IRT estimates and shifted the scale from (-4 -- 4) to (6 --14). As indicated in Figure 4.5, the mean of the depth of interaction equals 10.4 with the standard deviation of 1.802.

Insert Figure 4.5 in Appendix 4.B about Here

Instrument Diagnosis: Calculating Information Function

To estimate the information given each level of tie estimates, I took several steps: 1) calculated the mean of observed responses across these 10 items given each tie; 2) took the log of the mean to indicate the observed depth of interaction; 3) subtracted the estimates from the observed depth of interaction, then used the absolute values of these differences as the estimated measurement errors; and 4) took the inverse of the squared measurement errors to get the information (Lord, 1980).

I then plotted the information against the estimates of depth of interaction to evaluate the precision of this instrument. Figure 4.6 indicates that the survey instrument does not perform equally well across the continuum of the propensity of interaction. The survey instrument provides fairly sufficient information to inform the estimates of depth levels between 9.5 and 11.5, whereas it may perform less effectively to inform interaction levels lower than 9 or higher than 12.

Insert Figure 4.6 in Appendix 4.B about Here

Differential Item Function (DIF)

Table 4.10 includes the DIF estimates for each item. Model-I illustrates how white teachers responded to each item differently from non-white teachers, given non-white teachers as the reference group. Model-II includes results of DIF parameter estimates due to gender

differences, with males as the reference group. I found that compared to non-white teachers who shared the same (or the mean) propensity of seeking help with teaching mathematics, white teachers were less likely to endorse item-1, item-10, and item-12. These three items are among the most common collaborative activities. Moreover, compared to counterparts of males, females are more likely to endorse item-5, the rarest collaborative tasks.

Insert Table 4.10 in Appendix 4.C about Here

Prediction Model Estimates

I included three predictors to demonstrate how covariates could be included in the measurement model. As shown in Table 4.11, all coefficients are in the unit of log-odds. At the tie level, same racial group of both being White or both being non-White increased the odds that teachers engaged in instructional collaboration, after controlling for item rareness and other covariates in the model (β_{01j} =0.604, t-ratio= 1.789, p-value<0.1). This provides some tentative evidence of the homophily effect that teachers interacted with others who were similar to themselves. Moreover, after accounting for same racial group effect at the tie level and other covariates in the model, being a White teacher significantly decreased the teacher's propensity of reaching out to other teachers for professional help. (γ_{002} =-1.002, t-ratio= 1.789, p-value<0.1).

Insert Table 4.11 in Appendix 4.C about Here

Discussion

The purpose of this paper is to develop a psychometrically sound measure of the depth of teacher interaction from network data collected by survey instruments. I first articulated network theories that informed a multilevel item response theory modeling accounting for the differences in collaborative tasks (items) and the nested data structure. I then demonstrated a way of assessing the precision of the whole instrument and DIF of each item with regard to gender and race group effects. Finally, I illustrated how to include covariates in the measurement model to explore interesting research questions.

The theoretical framework is under an ego-centric network structure. The unique characteristics of an ego-centric network that assume independence across egos and relative independence across ties within each ego's network make the assumptions of a generalized hierarchical linear model plausible. However, this is not the case in sociocentric network where all individuals are connected. Moreover, reciprocity is another issue that we do not need to consider for ego-centric network data because in this study we only collected the single direction data from ego to alters. But for directed data, the effect of two-direction ties (e.g., teacher A offered help to teacher B and at the same time teacher B offered help to teacher A) should be considered in the estimation process. Therefore, the model developed in this study does not apply to analyze sociometric or directed network data.

The Rasch model used in this study is designed for binary responses. But some researchers have developed instruments to collect network data with more than two response categories (Gallagher et al., 2008; 2009). For example, teachers can rate the frequency of their interactions with other colleagues on different collaborative tasks on a five-point scale (0= "not at all", 1="once or twice this year", 2="monthly", 3="weekly", and 4="daily"). In this case, the

Rasch model is no longer appropriate to estimate the latent trait of the depth of interaction; rather, researchers need to employ polytomous models, such as Partial Credit Model (Masters, 1982) or Graded Response Model (Samejima, 1996), or Rating Scale model (e.g., Andrich, 1978a, 1978b; Andersen, 1997).

Moreover, this instrument and the estimates of depth of interactions may not be appropriate for some research purposes. For example, the instrument used to collect social network data in this study would not be very effective when researchers intend to identify one or two teachers who are most connected or least connected to colleagues for further case studies, because the instrument has large standard errors at the two ends of the depth distribution, as indicated in the information function. Moreover, as indicated by the information function, the instrument is effective to indicate a very narrow range of estimates of the latent trait around the level of depth 11. If researchers are interested in setting the cut-off point at this level, this instrument is well structured. But if researchers are interested in precisely measuring a wider range of the depth of interactions by flattening the distribution of information function, they can add some common (easy) items and some rare (difficult) items. Or if researchers are interested in setting the cut-off point at another level of depth of interactions, such as the level of 9, they can add more items with lower difficulty estimates; or if they want to set the cut-off point at the level of 12, they adding more items with higher difficulty estimates. In other words, with information function, researchers can restructure the instrument to serve specific purposes.

Researchers also need to be aware that the choice of reference item may have an impact on the interpretation of the log-odds for a particular tie, although the estimates of fixed effects for tie-level and ego-level predictor variable, the predicted values of depth of interactions, and

the rank order and distance between item parameters should not affected by which item is chosen as the reference item (Pastor, 2003).

The significant DIF racial effect on item-1, item-10, and item-12 and the significant DIF gender effect on item-5 deserve our attention. Traditionally, items that exhibit DIF are treated as biased items and normally deleted from the next test administration. Rather than simply exclude DIF items, recently, a consensus has emerged that DIF items indicate unexpected group effects and deserve further analysis, either controlling DIF effects in ability calibration or exploratory estimation, or collecting extra data to understand why DIF are present on these items.

Besides directly including predictors in the measurement model, researchers can also incorporate the estimates of the depth of interaction into traditional social network analysis models that Frank et al. used in his several studies. Frank has proposed two prediction models: social selection and social influence. The term *social selection* describes the process by which teachers seek resources and other possessors distribute these resources through communication (e.g. Banks & Carley, 1996; Snijders, 1996). As we mentioned previously, teachers may interact with each other for reasons related to either homophily or hetephily. Rather than having the measurement model at level one, we can directly use the estimate of the depth of interaction *j* at time *t* as the dependent variable and use other predictors, such as taught the same grade at time *t* and the difference in prior instructional expertise at *t-1* as predictors, as indicated in equation (4.10)

Depth of interaction $t_i = \pi 0 + \pi_1$ taught the same grade t_i

+
$$\pi_2$$
 difference in prior instructional expertise $_{t-1\,j}$ +...+ e_j (4.10)

Where in equation (4.10) π_0 represents the intercept. π_1 and π_2 represents the coefficients. e_j is the error term.

The term *social influence* refers to the process by which social actors change their behaviors or other sentiments as a result of exposure to new information or resources, influenced by others with whom they interact (e.g. Abelson & Bernstein, 1976; Burt, 1982). The extent to which the teacher is exposed to colleagues' influence is a function of the depth of their interactions, and the available expertise of colleagues, and the ability of colleagues to convey expertise (Frank et al., 2004), as indicated by equation (4.11)

Exposure to colleagues' expertise_i =

 $\sum_{\substack{i'=1,\\i'\neq i}}^{n_{i'}} (Depth of interaction_{ii'}) \times (Providers' prior expertise_{i'}) \times (Providers' ability to help_{i'})$

(11)

Where in equation (4.11) *i* indicates the ego, *i* ' indicate the alter. $n_{i'}$ indicates the number of alters who had helped ego with teaching mathematics.

The exposure variable can be then included in social influence models to estimate the extent to which exposure to colleagues' advices would change the teacher's own instructional practices, after controlling for teacher i's own prior instructional practices and other covariates, such as years taught mathematics, as illustrated in equation (4.12).

Instructional practices in 2009 $_i = \beta_0 + \beta_1$ Prior instructional practices (in 2008) $_i$

 $+\beta_2$ **Exposure** to colleagues' advices and help_i

+
$$\beta_3$$
 Years taught mathematics_i +...+ e_i . (4.12)

Conclusion

This paper adds methodological value to the use of social network data in educational studies by developing multilevel Rasch models to estimate the propensity of endorsing a relation between two teachers on matters of teaching mathematics and by developing methods to gauge the psychometric properties of the instrument. This multilevel Rasch model has transformed the binary scale of social network measure to a continuum and thus this IRT metric expands the family of social network analysis because more statistical models can employ a measure with a normal distribution than with a Bernoulli distribution. Furthermore, because IRT models can equate multiple survey instruments and put different estimates for the same tie on the same scale, this IRT metric surpasses the traditional methods (mean or sum of discrete responses) when modeling change in depth of interactions over time (Seltzer, Frank, & Bryk, 1994) or compare the depth of interactions across ties using equivalent survey instruments. This paper also demonstrates how to extend the multilevel measurement model to an explanatory model to explore multilevel research questions that we often encounter in educational settings. This combination of measurement and structural models may merit a better estimation of the true relationship between a predictor and the latent trait (Pastor, 2003; Maier, 2001; 2002).

Moreover, this study estimates the item information, DIF, and the information function of the whole instrument. This information can help researchers wisely use the survey instrument to inform further data analysis and the interpretations of results. Besides understanding what the instrument can precisely measure, researchers should be aware of the caveats of applying the data to answer practical questions, such as the differential functioning of items given subgroups of the sample and the standard errors associated with the estimates of latent trait. This information is also valuable for researchers to redesign the survey instrument in ways as

recommended in the discussion session. To add more value to use social network data in educational research, future studies are encouraged to develop models and methods to overcome the limitations of multilevel Rasch model by accounting for dependence and reciprocity in sociometric data and by analyzing categorical responses with more than two categories. **APPENDICES**

APPENDIX 4.A: INSTRUMENT

2. During this school year (including last summer), to whom have you turned for advice or information about teaching mathematics? Please write full first and last names (if known), and give a brief description of that person's role or position.

Name:	
Role:	

3.What type(s) of advice or information do you seek from this person? Please check all options that apply.

- o Doing mathematics problems together with discussion of different solution strategies
- Discussing why some students didn't learn as expected in a lesson in order to plan for future instruction
- Analyzing examples of student work in order to adjust instruction
- Analyzing examples of student work to understand the different ways that students solve problems.
- Discussing how to make use of student solution strategies in whole class mathematical discussions
- Discussing pacing
- Discussing what materials to use for a lesson
- After a lesson, sharing whether students "got it"
- Sharing materials or activities
- Analyzing student work to see if students "got it"
- Sharing materials or activities
- Analyzing student work to see if students" got it"
- Updating one another on a student or students' progress in mathematics
- Others (please specify)_____

APPENDIX 4.B: FIGURES

Figure 4.1 Egocentric Network Structure



Note: Hexagon represents egos. Oval represent alters. Solid line represents relations from egos to alters.

Figure 4.2 Egocentric Network Data Structure



Note: Items are included in rectangles. Alters are included in ovals. The ego is included in Hexagon. Dotted lines indicate associations between items and ties, while solid lines indicate ties.
Figure 4.3 The Ten Largest Eigenvalues in Order of Size

Size of the Eigenvalue



Note: The y-axis indicates the size of the eigenvalue, while the x-axis indicates the ranking of the 10 largest eigenvalues.

Figure 4.4 Graphical Comparison of Two-and One- Parameter model

(a) ICC in two paramter model: From top to bottom and from left to right: item-1, item -2, item-3, item-4, item-5, item-6, item-7, item-8, item-9, item-10, item-11, and item-12



(b) ICC in one-parameter model: From top to bottom and from left to right: item1, item 2, item 3, item 4, item 5, item 6, item 7, item 9, item 10, and item 12





Figure 4.5 The Distribution of Propensity of the Interactions

The Depth of Interactions



Figure 4.6 The Distribution of Information against the Propensity of the Interactions

APPENDIX 4.C: TABLES

Table 4.1 Item Descriptions

Itom	Itom I abol
Itelli	Itelli Lauei
Number	
Item-1	Doing mathematics problems together with discussion of different solution
	strategies
Item-2	Discussing different ways students are likely to solve tasks
Item-3	Discussing why some students didn't learn as expected in a lesson in order to
	plan for future instruction
Item-4	Analyzing examples of student work in order to adjust instruction
Item-5	Analyzing examples of student work to understand the different ways that
	students solve problems
Item-6	Discussing how to make use of student solution strategies in whole class
	mathematical discussions
Item-7	Discussing pacing
Item-8	Discussing what materials to use for a lesson
Item-9	After a lesson, sharing whether students "got it"
Item-10	Sharing materials or activities
Item-11	Analyzing student work to see if students "got it"
Item-12	Updating one another on a student or students' progress in mathematics

Characteristics	Mean
Percentage of race and ethnicity:	61.735%
White	
Black	28.718%
Asian	1.020%
Hispanic	5.612%
Latino	1.531%
Native American	2.551%
Percentage of female	68.205%
Percentage of teachers who held full teaching certification	89.189%
	5.053
Years of working experience in this school	(6.521)
	8.932
Years of teaching experience in mathematics	(9.122)

Table 4.2 Ego's Characteristics in the 2008-09 School Year

Note: Standard deviations are included in the parentheses.

Table 4.3 Te	etrachoric	Correla	ation l	Matrix
--------------	------------	---------	---------	--------

	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6	Item-7	Item-8	Item-9	Item-10	Item-11	Item-12
Item-1	1.000											
Item-2	0.695	1.000										
Item-3	0.597	0.518	1.000									
Item-4	0.514	0.628	0.488	1.000								
Item-5	0.632	0.696	0.673	0.879	1.000							
Item-6	0.648	0.637	0.576	0.854	0.844	1.000						
Item-7	0.686	0.770	0.510	0.648	0.726	0.640	1.000					
Item-8	0.219	0.162	0.480	0.285	0.308	0.277	0.302	1.000				
Item-9	0.554	0.525	0.558	0.398	0.432	0.486	0.483	0.482	1.000			
Item-10	0.662	0.561	0.639	0.575	0.557	0.699	0.547	0.425	0.624	1.000		
Item-11	0.347	0.262	0.453	0.396	0.362	0.285	0.246	0.444	0.679	0.429	1.000	
Item-12	0.490	0.391	0.591	0.523	0.567	0.541	0.574	0.325	0.548	0.719	0.396	1.000

	Error Variance	R-Square
Item-1	0.448	0.552
Item-2	0.383	0.617
Item-3	0.501	0.499
Item-4	0.261	0.739
Item-5	0.146	0.854
Item-6	0.281	0.719
Item-7	0.348	0.652
Item-8	0.800	0.200
Item-9	0.526	0.474
Item-10	0.302	0.698
Item-11	0.742	0.258
Item-12	0.458	0.542

Table 4.4 Error Variances and R-square by Items

	Two-Factor Model	One-Factor Model ^(a)
Goodness of Fit Index (GFI)	0.644	0.671
Bentler & Bonett's (1980) NFI	0.67	0.739
Chi-Square	1930.54	1296.923
Chi-Square DF	53	35
Standardized Root Mean Square Residual (SRMR)	0.094	0.086
RMSEA Estimate	0.259	0.261
Bentler's Comparative Fit Index (CFI)	0.675	0.744

Table 4.5 Compare Goodness of Fit between Two-Factor Model and One-Factor Model Using CFA

Note: ^(a) One-factor model is after removing item-8 and item-11. The cut-off values of goodness fit for GFI and NFI indices should be close to 1.

Table 4.6 Factor Loadings

Factor Loadings
0.782
0.779
0.728
0.801
0.878
0.864
0.803
0.641
0.792
0.695

	Two-Parameter Model			Rasch Model	
	Biserial Correlation	Slope	Difficulty	Biserial Correlation	Difficulty
Item-1	0.736	2.088 (0.253)	-0.327 (0.071)	0.756	-0.229 (0.052)
Item-2	0.719	2.15 (0.289)	-0.521 (0.074)	0.755	-0.571 (0.053)
Item-3	0.732	2.027	-0.668	0.702	-0.804 (0.053)
Item-4	0.775	2.683 (0.343)	-0.193	0.787	0.002
Item-5	0.835	3.315	-0.003	0.858	0.39
Item-6	0.816	3.125	-0.136	0.85	0.118
Item-7	0.747	2.225 (0.269)	-0.124	0.774	0.128 (0.052)
Item-8	0.418	(0.205) 0.892 (0.145)	-1.66		(0.052)
Item-9	0.691	(0.143) 1.625 (0.237)	(0.2+3) -0.99 (0.111)	0.616	-1.194 (0.054)
Item-10	0.786	2.232 (0.273)	-0.393 (0.071)	0.775	-0.355 (0.052)
Item-11	0.485	1.01 (0.157)	-0.966 (0.147)		
Item-12	0.671	1.645 (0.194)	-0.339 (0.081)	0.664	-0.21 (0.052)
-2 Log likelihood	7	153.4271		5993.	9442

Table 4.7 One- and Two-Parameter Models

Notes: Standard errors of estimates are included in the parentheses.

	The Level 2 Tie Le	evel	The Level 3 Ego Level	
	Estimates of Variance Components	p-value	Estimates of Variance Components	p-value
Item-1	0.002	>.500	0.359	0.191
Item-2	0.003	>.500	0.329	>.500
Item-3	0.003	>.500	0.305	0.470
Item-4	0.002	>.500	0.369	0.428
Item-5	0.002	>.500	0.214	>.500
Item-6	0.002	>.500	0.183	>.500
Item-7	0.002	>.500	0.314	0.027
Item-8	0.014	0.004	14.011	0.000
Item-9	0.003	>.500	0.574	0.003
Item-10	0.002	>.500	0.204	>.500
Item-11	0.006	>.500	1.467	0.000

Table 4.8 Item Goodness-Fit Indices under Multilevel Framework

Table 4.9 Item Rareness Estimates

	Item Description	Coefficients	Rank
Itom 12	Updating one another on a student or students' progress in	-0.386	1
110111-12	mathematics	(0.154)	1
Itom 3	Discussing why some students didn't learn as expected in a	0.101	2
nem-5	lesson in order to plan for future instruction	(0.139)	2
Itom 2	Discussing different ways students are likely to solve tasks	0.395	3
110111-2	Discussing unreferit ways students are likely to solve tasks	(0.145)	5
Item_0	After a lesson sharing whether students "got it"	0.667	Δ
Item-7	After a lesson, sharing whether students got it	(0.151)	-
Item_1	Doing mathematics problems together with discussion of	0.826	5
10111-1	different solution strategies	(0.126)	5
Item_10	Sharing materials or activities	0.851	6
10111-10	Sharing materials of activities	(0.130)	0
Item_/	Analyzing examples of student work in order to adjust	1.121	7
110111-4	instruction	(0.114)	/
Item_6	Discussing how to make use of student solution strategies in	1.269	8
num-0	whole class mathematical discussions	(0.138)	0
Item_7	Discussing pacing	1.282	9
10111-7	Discussing pueling	(0.130))
Item_5	Analyzing examples of student work to understand the	1.620	10
10111-5	different ways that students solve problems	(0.129)	10

Note: Standard errors are included in the parentheses. Items are presented in the order of rareness level from the most common to the rarest one.

	Model-I White	Model_II Female
	Coefficients	Coefficients
DIE an Itam 1	-1.264**	-0.425
DIF on Item-1	(0.334)	(0.369)
DIE on Item 2	0.319	0.850
Diff on Rem-2	(0.381)	(0.443)
DIF on Item_3	-0.536	0.370
Dir on item-5	(0.351)	(0.388)
DIF on Item-A	0.563	0.607329
Dir on nem-4	(0.454)	(0.456)
DIF on Item-5	0.318	1.046**
Dir on item 5	(0.409)	(0.496)
DIF on Item-6	0.300	0.222
Dir on nem-o	(0.458)	(0.521)
DIF on Item-7	0.011	0.500
Dir on item-7	(0.374)	(0.379)
DIF on Item-9	-0.645	-0.293
	(0.419)	(0.459)
DIF on Item-10	-1.034*	-0.101
Dir on nem-10	(0.354)	(0.407)
DIF on Item-12	-0.749*	-0.211
Dir on item 12	(0.329)	(0.380)
Deviance	12410.069	12438.86
	(df=22)	(df=22)

Table 4.10 Differential Item Function (DIF) Parameter Estimates

Notes: Ego N=167, Tie N=438, and Item N=4380 *p-value≤0.05; **p-value≤0.01

Table 4.11 Fixed Effects of Prediction Model

Variables	Coefficients
<i>Tie level (N=252)</i>	
Same regist group of both being white or both being non white	0.604
Same factal group of both being white of both being holf-white	(0.338)
Ego Level (N=143)	
Ego's famela	0.176
Ego's tennale	(0.377)
Fao's white	-1.002**
Lg0 5 winte	(0.348)

Notes: Item level N=2520 and including nine item dummy variables in the model. Standard Errors are included in the parentheses. Notes: $p-value \le 0.05$; $p-value \le 0.01$

- Abelson, R., A. Bernstein (1976). A computer simulation model of com-munity referendum controversies. *Public Opinion Quart*. 27, 93-122.
- Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, 22(1), 47-76.
- Anderson, E. B. (1997). The rating scale model. In W. v. d. L. R. K. H. (Eds.) (Ed.), *Handbook* of modern item response theory (pp. 67-84). New York: Springer.
- Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581-594.
- Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques* (Second ed.). NY: New York Marcel Dekker, Inc.
- Banks, D., & Carley, K. M. (1996). Models of Social Network Evolution. Journal of Mathematical Sociology, 21(1-2), 173-196.
- Briggs, D. C. (2008). Using explanatory item response models to analyze group differences in science achievement. *Applied Measurement In Education*, 21, 89-118.
- Burt, R. S. (1982). *Toward a structural theory of action: Network models of social structure, perception, and action.* New Work, NY: Academic Press.
- Christofersson, A. (1975). Factor analysis of dichotomized variables. Psychometrika, 40, 5-32
- Cobb, P., McClain, K., Lamberg, T. d. S., & Dean, C. (2003). Situating teaches' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*(6), 13-24.
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.
- Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis*, 30(3), 203-235.
- Cross, J. E., Dickman, E., Newman-Gonchar, R., & Fagan, J. M. (2009). Using mixed-method design and network analysis of measure development of interagency collaboration. *American Journal of Evaluation*, 30(3), 310-329.

De Boeck, P. (2008). Random item IRT models. Psychometrika, 73(4), 533-559.

- De Boeck, P., & Wilson, M. (Eds.) (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.
- Fox, J.-P. (2007). Multilevel IRT modeling in practice with the package mlirt. *Journal of Statistical Software*, 20(5), 1-16.
- Frank, K. A. and Zhao, Y. (2005). *Subgroups as a Meso-Level Entity in the Social Organization of Schools*. Chapter 10, pages 279-318. Book honoring Charles Bidwell's retirement, edited by Larry Hedges and Barbara Schneider. New York: Sage publications
- Frank, K. A., Zhao, Y., & Borman, (2004). Social Capital and the Diffusion of Innovations within Organizations: Application to the Implementation of Computer Technology in Schools. Sociology of Education, 77, 148-171.
- Frank, K., A., & Yasumoto, J. Y. (1998). Linking action to social structure within a system: Social capital within and between subgroups. *American Journal of Sociology*, 104(3), 642-686.
- Frank, K.A., Kim, C., & Belman, D. (2010). Utility Theory, Social Networks, and Teacher Decision Making. Pages 223-242 in Alan J. Daly editor. *Social Network Theory and Educational Change*. Cambridge: Harvard University Press.
- Gallagher, H. A., Woodworth, K. R., Bosetti, K. R., Cassidy, L., McCaffrey, T., Yee, K., et al. (2010). National Evaluation of Writing Project Professionall Development Year 4 Report: SRI International.
- Garrison, A., & Smith, T. (2011). *Investigating School and Individual Factors that Influence Teachers' Learning Opportunities through Interactions*, The annual meeting of American Educational Research Association. LA: New Orleans.
- Hu & Bentler (1999). Cutoff criteria for fit indexes in covariance structure analysis: Coventional criteria versus new alternatives, *Structural Equation Modeling*, 6(1), 1-55.
- Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, *38*(1), 79-93.
- Kamata, A., Bauer, D.J. & Miyazaki, Y. (2008). Multilevel measurement modeling. In A.A. O'Connell & D.B. McCoach (Eds.) *Multilevel Modeling of Educational Data* (pp. 345-388). Charlotte, NC: Information Age Publishing.
- Kelley, C., & Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational Administration Quarterly*, 33(4), 474-505.
- Lord, F. M. (Ed.). (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

- Maier, K. S. (2001). A rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, 26(3), 307-330.
- Maier, K. S. (2002). Modeling incomplete scaled questionnaire data with a partial credit hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, 27(3), 272-289
- Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174
- Meulders, M., & Xie, Y. (2004). Person-by-item predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 213-240). New York: Springer.
- Muth én, B. (1977). Statistical methodology for structural equation models involving latent variables with dichotomous indicators. Unpublished doctoral dissertation, Uppsala University.
- Osgood, D. W., McMorris, B. J., & Potenza, M. T. (2002). Analyzing multiple item measures of crime and deviance 1: Item response theory scaling. *Journal of Quantitative Criminology*, *18*(3), 267-296.
- Pastor, D. A. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *Applied Measurement In Education*, *16*(3), 223-243.
- Raudenbush S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd edition). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Johnson, C., & Sampson, R. J. (2003). A multivariate, multilevel Rasch model with application to self-reported criminal behavior. *Sociological Methodology*, 33, 169-211.
- Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *Journal of Computational and Graphical Statistics*, 9(1), 141-157.
- Reckase, M. D. (2009). Multidimensional Item Response Theory. New York, NY: Springer.
- Samejima, F. (1996). Evaluation of mathematical models for ordered polychotomous responses. *Behaviormetrika*, 23(1), 17-35
- Seltzer, M. H., Frank, K., A., & Bryk, A. S. (1994). The metric matters: The sensitivity of conclusions about growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis*, 16(1), 41-49.
- Snijders, T. (1996). Stochastic actor-oriented models for network change. *Journal of Math. Sociology*, 2 (1-2) 149-172.

- Supovitz, J., & Weinbaum, E. H. (2008). Reform Implementation Revisited. In J. Supovitz & E. H. Weinbaum (Eds.), *The implementation gap: Understanding reform in high schools*. New York: Teachers College Press.
- Wellman, B., & Frank, K., A. (2001). Network capital in a multi-level world: Getting support from personal communities. In R. B. a. K. C. Nan Lin (Ed.), *Social Capital: Theory and Research*. Chicago: Aldine De Gruyter.
- Yang, M. (1998). Increasing the Efficiency in Estimating Multilevel Bernoulli Models, unpublished Ph.D.dissertation, Michigan State University, Department of Counseling, Educational Psychology, and Special Education.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R.D. (2003). *BILOG-MG 3 for Windows: Multiple-group IRT analysis and test maintenance for binary items [Computer software].* Lincolnwood, IL: Scientific Software International, Inc.

CHAPTER 5: SUMMARY AND CONCLUSIONS OF THIS DISSERTATION

This dissertation details the results of the investigation of how school intra-organizational mechanisms mediate the implementation of educational interventions on teaching and learning. The findings are based upon analyzing data on teacher professional networks, instructional practices, school leadership, professional development programs, and individual background characteristics from three large-scale longitudinal studies. The unique databases of longitudinally connecting professional networks with individual practices have allowed me to rigorously examine the complex mechanism and the interplay of school formal and informal supports for instructional changes. Having taken evidence from all these three sub-studies, this dissertation provides a set of suggestions for policy makers and school leaders about how to orchestrate individual and organizational support for instructional improvement, including designing effective professional development programs, developing principal instructional leadership, and building teacher professional communities.

First, in Chapter 2, the first sub-study highlights a paradigm shift of thinking on how professional development programs should work. The messages promoted in professional development programs may be adopted by individual teachers and then channeled to other teachers through professional interactions. Beyond examining the direct impact, school leaders or policy makers should recognize the ways that professional developments can indirectly affect teachers' instructional practices. To promote both direct and indirect effects, this dissertation reveals several effective PD features, including 1) longer contact hours; 2) content foci on subject pedagogical knowledge and strategies and skills to collaborate with colleagues; and 3) collective and interactive formats, such as in-classroom coaching or mentoring, actively discussing classroom implementation with co-participants or PD providers, analyzing students'

151

work with other teachers, and receiving constructive feedback on their classroom teaching. These effective PD programs greatly promote not only PD participants' instructional behaviors, but also knowledge diffusion among teachers.

Second, principal leadership is a catalyst for implementing external reforms. This dissertation suggests two key ideas of developing principals' instructional leadership. The first is to a strategic focus on improving the general practices of teaching and targeting leadership thinking regarding the technical layer of reform, such as setting educational goals, collecting instructional materials, and developing assessments for teaching and learning. This idea stems from the findings in Chapter 3 and bases on the theory of shared instructional leadership. The other key idea is grounded in the efforts to build collaborative norms and trust relationships across the school community by allocating scheduled time for teacher collaboration and wisely distributing expert teachers in the school organization to promote knowledge diffusion. This idea capitalizes on the results in Chapter 2 which is based on the theory that schools, as social organizations, must be improved through promoting social learning (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010).

Third, to support continuous improvement, individual teachers must be supported by a coherent and collaborative professional teaching community. This kind of professional teaching community can be both social and knowledge resources for improvement. The second sub-study in Chapter 3 reveals social support and normative pressure from peers on the teachers' individual improvement of specific pedagogical practices, which constitutes the inner layer of technical core of instruction. The first sub-study supports knowledge diffusion among teachers and the belief that teachers can be the most successful teachers when engaging with other teachers. Developing the strength in such professional teaching supports in the local school community

152

can be essential to sustain the hard work toward instructional improvement. Moreover, teachers who proactively engage in collaboration and offering help to others can be teacher leaders potentially the greatest resource for educational reforms.

In conclusions, although these essential supports are not newly discovered, the unique contribution of this dissertation is to explain and predict the interwoven relationships between individual supports (e.g., professional development) and organizational mechanisms (e.g., leadership and collegial community). This dissertation also provides discernible pieces of evidence on the processes that effects of external interventions on individual outcomes take place in schools and are mediated by the intra-organizational networks among school staff. Every policy agenda that aims to bring sustainable and profound improvements of instructional practices should be grounded on the view that teachers live in a social and learning organization. I would like to close this dissertation by quoting Bryk et al.'s statement that to inform policy and practice of improving teaching and learning, research studies "most do more than just 'tell the facts'"; rather, "we must seek to understand , and we must also ask why" (2010, p.222). I hope this dissertation has contributed to better exploring and understanding practice and policy of improving teaching and learning.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: The University of Chicago Press.