How Mentors, Colleagues, Relational Trust and Collective Responsibility Influence Outcomes for Early Career Teachers and Their Students
Peter Youngs, Ken Frank, Yeow Meng Thum, and Mark Low. Michigan State University
March 16, 2008

Schools and districts face several challenges with regard to the instructional practices, effectiveness, commitment, and retention of beginning elementary and middle school teachers. Many new teachers lack the subject matter knowledge needed to help students achieve at high levels in literacy, mathematics, and other subjects (e.g., Pressley, Duke, & Boling, 2004; Hill, Rowan, & Ball, 2005). Further, many struggle to enact student-centered practices and to promote students’ higher-order skills in these subjects. With increased attention to student performance under No Child Left Behind, (NCLB) it is not sufficient for schools and districts to provide novices with opportunities for professional growth; they also need to attend to their effectiveness in fostering student learning. In addition, with new requirements under NCLB and the Individuals with Disabilities Education Act (IDEA) regarding the inclusion of students with disabilities in state testing, schools and districts must also attend to the practices and impact of beginning special education teachers.

In terms of retention, research indicates that close to 30 percent of first-year general education and special education teachers in 1999-2000 either left teaching at the end of the year or migrated to other schools or districts (Smith & Ingersoll, 2004). Further, research has found teacher attrition in the first five years among general and special education teachers to be 30 percent or greater (Henke, Chen, & Geis, 2000). These characteristics of teacher quality and instructional practice in the U.S. contribute to low levels of student performance on the National Assessment of Educational Progress and on international comparisons (Graham & Perin, 2007; Kloosterman & Walcott, 2007; National Center for Education Statistics, 2003; Schmidt et al., 2001).

In response to these concerns, many districts and states have implemented induction programs for new teachers (Education Week, 2008) and 8 out of 10 novices in 1999-2000 reported participating in an induction program (Smith & Ingersoll, 2004). Designed to help beginning teachers acclimate to their schools and districts as well as the profession, these programs typically feature mentoring, orientation sessions, workshops and seminars, and opportunities for novices to collaborate with colleagues. Further, many districts offer professional development for new and experienced teachers that addresses curriculum, instruction, and assessment in reading, writing, and math. With increased attention to student performance under the federal No Child Left Behind legislation, districts are seeking ways to assess the impact of these activities on early career teachers’ instructional practices and retention, as well as student achievement (see, for example, Desimone et al., 2002).
But how do we explain the effects of mentors, colleagues, and professional development on beginning teachers? In research on a variety of topics, one of the most striking and robust results is that across studies 66 to 90 percent of the variation in student or teacher outcomes is within schools (Frank, 1998; Frank & Zhao, 2005; Rowan, Camburn, & Correnti, 2004). Characteristics of students and teachers, such as socio-economic background, gender, race, education, and experience explain a portion of this variance. But what accounts for the remaining variation? In this paper, we describe a three-year research study, known as the Michigan Indiana Early Career Teacher (MIECT) Study. The purpose of the MIECT Study is to investigate how mentors, colleagues, and professional development contribute to variations in teachers’ instructional practices, commitment levels and retention decisions; and their effects on student achievement. The theory underlying our study posits that there can be considerable tensions between a) new teachers’ beliefs about effective teaching (i.e., their individualistic motivation) and b) the beliefs or expectations (regarding what they should do) of others who make up the social system of their schools (Youngs et al., under review). Drawing on economics, we explain how utility functions can be used to express this fundamental tension and to measure the impact of mentoring, colleagues, and professional development on novices’ instructional practices and student achievement (Akerlof & Kranton, 2002; Frank et al., in press).

The first section of this paper reviews the research literature on the mentoring and induction of beginning teachers in general and special education. In the second section, we present the theory that underlies our research design and approach to data analysis. Third, we discuss our samples in 2006-07 and 2007-08, our research methods, and several hypotheses that we are investigating in this study. The fourth section describes our plans for analyzing effects of mentors, colleagues, and professional development on novices’ instructional practices and student learning gains. Finally, we conclude by describing next steps for the MIECT Study and considering the implications of our interim findings for future research on induction.

I. Research on Mentoring and Induction

There has been extensive writing about mentoring and induction over the past two decades (for reviews of the research, see Little, 1990; Gold, 1996; and Smith & Ingersoll, 2004). Until recently, though, there had been few studies of induction that were well-grounded in theory, took account of school context and/or district policies, and included measures of key teacher outcomes. In the past ten years, several well-designed qualitative and large-scale quantitative studies have been completed, and these studies represent important advances in research on beginning teachers. In this section, we describe
findings from these studies and consider their contributions to what is known about mentoring and induction as well as the limitations of current research.

A number of recent qualitative studies have examined how district policies and school organizational conditions influence beginning teachers’ experiences. For example, Grossman and Thompson (2004) investigated opportunities to learn to teach language arts that were provided to three first-year secondary teachers in two school districts in the state of Washington. They reported that the instructional support available to these teachers seemed to be shaped by the degree of alignment between themselves and their mentors, mentor characteristics, and the extent of collaboration in their departments. They also found that district and state policies related to curriculum, assessment, and professional development seemed to influence novices’ learning opportunities and practices (Grossman & Thompson, 2004). In a study of two first-year elementary teachers in two California districts, Achinstein, Ogawa, and Speiglman considered how teacher characteristics, alignment with mentors and colleagues, collaboration, and expectations affected their induction experiences. The authors reported that these factors, along with teacher preparation and district hiring policies, seemed to be associated with variations in the two teachers’ experiences (Achinstein, Ogawa, & Speiglman, 2004).

In a third study, Youngs (2007) examined connections between district policies and the nature of the assistance provided to first- and second-year teachers in two school districts in Connecticut. He found that district policy related to mentor selection, mentor assignment, and professional development seemed related to variations in beginning teachers’ induction experiences across the two districts (Youngs, 2007). In particular, teachers in one district consistently had greater opportunities to acquire curricular knowledge, plan instruction, and reflect on practice. Finally, Johnson and colleagues (2004) investigated the effects of school culture on new teachers’ experiences. They reported that schools with integrated cultures were characterized by high levels of collaboration between novice and experienced teachers. In these schools, there were frequent opportunities for early career teachers to talk with their mentors and other veteran colleagues about curriculum, instruction, and assessment (Johnson & PNGT, 2004; Kardos et al., 2001).

With regard to special education, relatively few empirical studies have focused on novice special educators. For these teachers, we do not have a clear picture of how mentoring, other induction activities, or relationships with colleagues influence key outcomes. While we could apply the findings from studies of induction in general education to special education, to do so would overlook the differences in how the two categories of teachers are socialized into schools. One noteworthy study on first-year teachers was conducted by Whitaker (2000), who surveyed over 150 first-year special educators about their relationships with their mentors, finding that when relationships with mentors are informal and more personal, teachers are more likely to intend to stay in teaching. In addition, it appears that for beginning
special educators, mentors in special education can offer unique skills and content knowledge that general education teachers cannot, as well as providing emotional support (Whitaker, 2000; Kueker & Haensly, 1991). If a novice’s mentor is in special education, she or he is likely to be one of the few individuals in the beginning teacher’s school who has gone through a similar work experience. Looking beyond the relationship between the mentor and mentee, Billingsley, Carlson, and Klein (2004) found that 89 percent of new special education teachers indicated that “informal help from other colleagues” was helpful to a moderate or great extent.

Along with these qualitative studies, researchers have employed large-scale data sets to consider how mentoring and other factors impact the retention of beginning teachers (see also Glazerman et al., under review). Using nationally representative data from the 1999-2000 Schools and Staffing Survey (SASS), Smith and Ingersoll (2004) investigated the influence of mentoring and other induction activities on first-year teacher attrition and migration while controlling for the possible effects of other factors. They reported that for general education teachers, having a mentor in one’s field reduced the risk of leaving teaching by about 30 percent, but had little effect on teacher migration (i.e., moving from one’s school of origin to another school). In addition, collaborating regularly with colleagues on instruction reduced the risk of leaving teaching by 43 percent and lowered the risk of migration by 25 percent. In a large-scale study of first-year teachers in Chicago, Kapadia, Coca, and Easton (2007) found that several factors predicted positive classroom experiences and commitment to the profession among beginning general education teachers. These factors included the quality, intensity, and level of mentoring and support, student composition (i.e., number of students with behavioral problems), a welcoming faculty, and school leadership (Kapadia, Coca, & Easton, 2007).

While these studies represent an important advance in research on induction, they have several limitations. First, few of these studies followed new teachers over time to examine how mentoring, other induction activities, and professional development affect their instructional practices, effectiveness, commitment, and retention over two or more years. Second, these studies rarely investigated the instructional practices of mentors, colleagues, and other key individuals; their expectations for new teachers; or the nature of their interactions with novices. Third, with some exceptions, few of these studies sufficiently controlled for the effects of student and teacher characteristics on various outcomes. Fourth, few studies have considered how the induction and socialization experiences of new special education teachers differ from those of their counterparts in general education. Finally, few of these studies account for the fact that schools are fundamentally social organizations characterized by social psychological processes (Bidwell, 2000; Frank, Zhao, & Borman, 2004; Portes, 1998). That is, individuals, induction activities, and policies can also place expectations on early career teachers or even exert pressures on them (Kennedy, 2005; Youngs et al., under review).
II. A Formal Model of Teacher Effort

As noted, there can be considerable tensions between a) beginning teachers’ beliefs about what they need to do to teach effectively and b) the beliefs or expectations (regarding what they should do) held by other individuals who define the social system of their schools. The first of these aspirations is associated with psychic rewards related to teaching (Bandura, 1977; Hargreaves, 1993; Lortie, 1975) while the latter involves the degree of social fit between oneself and colleagues (Bidwell, 2000; Zhao & Frank, 2003; McLaughlin & Talbert, 2001). The extent and resolution of these tensions can affect key outcomes for early career teachers, including their instructional practices, commitment levels, and retention decisions, and student learning. Drawing on economics, this fundamental tension can be expressed by a utility function. The utility for teacher $i$, given membership in a given context $C$, is

$$U_i(C) = f(\text{effort}, \text{payoffs of effort, expectations of others})$$  \hspace{1cm} (1)

Together with Figure 1, equation (1) helps to explain changes in teachers’ instructional practices and student achievement over time. As illustrated in Figure 1, a main goal for each teacher is to help students learn (i.e., acquire human capital), and the teacher’s contribution to learning is a function of the product of their effort and the results or payoffs associated with effort. Mentors, colleagues, and professional development activities can increase the payoffs of beginning teachers’ effort (i.e., their contribution to student learning) by helping them acquire subject matter and curricular knowledge, as well as instructional expertise.
At the same time, early career teachers must resolve tensions between their own beliefs about effective teaching (i.e., their individualistic motivation) and expectations placed on them (regarding their instructional practices) by other individuals, professional development activities, and policies. Figure 1 indicates that mentors and colleagues can potentially help novices make sense of the expectations placed on them and resolve these tensions. When significant tensions between a novice’s beliefs and others’ expectations remain unresolved over time, though, these tensions can reduce the teacher’s effectiveness (i.e., their contribution to student learning) and lower their commitment.

Drawing on research on adolescents by Akerlof and Kranton (2002) and Frank et al. (in press), we describe a way to model the tensions between new teachers’ own beliefs about effective teaching and others’ beliefs and expectations regarding their practice. Elaborating the expression in equation (1) above, the utility for teacher $i$, given membership in a given context $C$, is

$$U_i(C) = \rho \left[ e_i k_i - \frac{1}{2} e_i^2 \right] + (1 - \rho) \left[ -\frac{1}{2} (e_i - X(C))^2 \right].$$

In this expression, $C$ refers to relevant contexts for beginning teachers such as mentoring, colleagues, or professional development. Effort is represented by $e$, while $k_i$ represents the payoffs in, for instance, adolescent human capital associated with a teacher’s effort. In teaching math, for example, some elements in $k$ would include the teacher’s mathematical content knowledge, their selection of mathematical tasks, the instructional practices they use to implement those tasks, and the ability of their students to learn math. For instance, the greater a given teacher’s content knowledge and/or the greater their students’ ability, the greater the return on that teacher’s effort. In addition, teacher- and school-level factors such
as job manageability, collective responsibility, and collective efficacy mediate the payoffs of effort (Bryk & Schneider, 2002; Gold, 1996; Lee & Smith, 1996; Goddard, Hoy, & Hoy, 2004). (See Figure 2.)

![Figure 2 Social Context for Beginning Teachers II](image)

Figure 2  Social Context for Beginning Teachers II

According to equation (2), the contribution to relevant outcomes (e.g., teachers’ instructional practices, commitment levels, and student achievement) is a function of the product of effort ($e_i$) and the payoffs associated with effort ($k_i$). This contribution is then balanced off against effort ($-\frac{1}{2}e_i^2$) in the terms associated with ($\rho$). At the same time, the teacher must resolve their individualistic motivation to develop student human capital with the desire to conform with expectations of other individuals, policies, etc.: $(1-\rho)\left[-\frac{1}{2}(e_i - X(C))^2\right]$. In this part of the utility function, X refers to the expectations of those individuals who make up social context C (e.g., mentor, colleagues).

A novice teacher may feel little professional tension when her expectations for herself are consistent with group norms because she can produce human capital and conform by exerting effort on a given dimension. When her expectations are not aligned with group norms, though, she may experience significant tensions that can have negative effects on her instruction and effectiveness. For example, a teacher could enter the profession with the intention of implementing tasks of high cognitive demand regularly and she might possess the content and curricular knowledge to do so. However, colleagues
could convey to this teacher that her students are not capable of doing meaningful work on high-level tasks and that they expect her to focus primarily on teaching students algorithms and other forms of procedural knowledge. This tension could have a negative impact on the novice teacher’s effectiveness.

Expressing theory through a utility function has a distinct advantage for developing a theory of teacher effort and response to expectations within a social context. Utility functions represent a way of ordering preferences for different quantities of non-monetary goods (e.g., effort, support from mentors and colleagues, expectations of others, school characteristics). This ordering facilitates an interdisciplinary understanding of factors that influence novice teachers’ instructional practices, commitment levels, and retention decisions, and student achievement. In addition, this ordering operationalizes the assumption of rational action without reduction to monetary terms – utility is a function of one’s own beliefs about effective teaching and learning as well as the sociological value of conformity with others’ expectations.

In theorizing the effects of mentoring, colleagues, and professional development on teacher and student outcomes, it is important to note two related issues. First, an individual teacher’s instructional practice is actually a set of practices and, following equation (2), these practices are shaped by that teacher’s characteristics as well as their work context. In terms of the former, a teacher’s practices are informed by their professional backgrounds, years of experience, curricular knowledge, teaching expertise, and out-of-work responsibilities. With regard to the latter, as discussed above, a teacher’s practices can also be influenced by the practices and expectations of their mentor, colleagues, and others.

Second, teacher effort regarding instruction and student learning can be multidimensional in a number of ways. For example, elementary school teachers teach different subjects (e.g., math, reading, writing, science) to one or more groups of students. For their part, most middle school teachers teach different courses within one content area (e.g., algebra, geometry, pre-calculus within math) to several groups of students. These differences between elementary and middle school teachers are likely to result in varying demands on content knowledge as well as different conceptions of students and their abilities. In addition, within a given subject, teachers vary in the ways in which and the extent to which they balance basic skills and other forms of procedural knowledge with higher-order thinking, problem-solving, and completing tasks of high cognitive demand. Differences in the ways teachers handle these tensions are visible in the instructional practices that they utilize (e.g., Henningsen & Stein, 1997). Further, teachers vary in the ways in which and the extent to which they emphasize authority, moral issues, and relationships with students in their teaching (Bidwell, Frank, & Quiroz, 1997).

The fact that teacher effort with regard to instruction is often multidimensional in these ways can be expressed as $e_{id}$, representing the amount of effort teacher $i$ devotes to activity $d$ with $d$ connoting a) demands on content knowledge, b) relative emphasis on basic skills vs. higher-order skills, c) relative
emphasis on authority, moral issues, etc. The utility function in (2) can then be modified to represent effort in any particular area:

\[ U_{id}(C) = \rho_d \left[ e_{id} k_{id} - \frac{1}{2} e_{id}^2 \right] + (1 - \rho_d) \left[ -\frac{1}{2} \left( e_{id} - X_d(C) \right)^2 \right]. \] (3)

Although the expression for the utility in equation (3) may appear awkward involving squared terms and coefficients of 1/2, it yields simple expressions when maximized with respect to effort (known as the first order condition for effort). This can also be understood as the minimum amount of effort required to achieve a given utility (Deaton & Muellbauer, 1980). Therefore, assuming a budget constraint in the form of a fixed amount of time/effort, utility is maximized with respect to effort when

\[ e_{id} = \rho_d k_{id} + (1 - \rho_d) e_{i}(C) \] (4)

The implications of (4) are fairly straightforward. Assuming payoffs for effort based on \( k_i \), expectations of others as represented by \( e(C) \) and that teachers act based on their preferences, equation (4) expresses the relative importance of payoffs to effort (associated with \( \rho \)) and conformity to the social context (associated with \( 1 - \rho \)). We can then estimate parameters such as \( \rho \) and \( (1 - \rho) \) in a model with effort as the dependent variable.

In sum, in order to understand the effects of mentors, colleagues, and professional development on beginning teachers, the MIECT study employs utility functions to model the relationships between teachers’ efforts to teach based on their own beliefs about effective teaching and their efforts to conform to the expectations of others in their schools.

III. Research Methods

As noted, the MIECT Study is investigating associations between mentoring, collaboration with colleagues, and professional development and a) new elementary and middle school teachers’ instructional practices, b) their commitment levels and retention decisions, and c) their effects on student learning gains. For novice special education teachers, we are focusing on associations between these factors and their instructional practices, commitment levels, and retention decisions. The success of this study in meeting these objectives depends on the following methodological decisions: identifying district and early career teacher samples that maximally serve our goals, collecting survey and student achievement data over multiple years from beginning teachers and their mentors and colleagues, and selecting strategies for analyzing survey and student achievement data. The next two sections explicate the data collection methods and analysis strategies for the new teacher, mentor colleague, and student achievement data.
**District Sample.** This study included four Michigan school districts in 2006-07 (the first year of data collection) and was expanded to include six Michigan districts and five Indiana districts in 2007-08. In selecting the district sample, our goal in 2006 was to recruit medium-to-large districts in Michigan that a) served varying student populations with regard to race/ethnicity and socio-economic status and b) had significant numbers of early career teachers. Because of declining enrollments and tight fiscal budgets in Michigan, many districts that served large numbers of low-income and racial minority students did not meet the other criterion for inclusion in this study because they did not hire new teachers for the 2006-07 school year. In fact, many districts in Michigan, including a few in this study, laid off teachers at the end of 2005-06 and 2006-07, although most of these teachers were later rehired.

In 2006, we recruited four Michigan districts to participate in the study during the 2006-07 school year: Daus, Greenberg, Kaline, and Whitaker. In 2006-07, these districts ranged in size from about 9,500 students to more than 27,000 students. Further, in two of the districts, Whitaker and Kaline, about 30 to 40 percent of the students were eligible for free and reduced-price lunch while in the other two districts, Daus and Greenberg, about 50 to 65 percent of the students were eligible for free and reduced-price lunch. Finally, one of the districts, Whitaker, primarily served white students; about 80 percent of its enrollment was white in 2006-07. In contrast, in 2006-07, 50 percent of the students in Kaline were racial minorities as were 80 percent of the students in Greenberg. In Daus, almost 90 percent of the students in 2006-07 were classified as white, although this included a large percentage of immigrant students of Middle Eastern decent., many of whom were English language learners. Table 1 summarizes enrollment and demographic information from the Michigan Department of Education for these four districts in 2006-07.

<table>
<thead>
<tr>
<th>District</th>
<th>Total Enrollment</th>
<th>% Free/Reduced Lunch</th>
<th>% Non-white</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daus (MI)</td>
<td>19,055</td>
<td>51%</td>
<td>12%</td>
</tr>
<tr>
<td>Greenberg (MI)</td>
<td>27,066</td>
<td>64%</td>
<td>80%</td>
</tr>
<tr>
<td>Kaline (MI)</td>
<td>9,448</td>
<td>42%</td>
<td>50%</td>
</tr>
<tr>
<td>Whitaker (MI)</td>
<td>12,354</td>
<td>29%</td>
<td>19%</td>
</tr>
</tbody>
</table>

Based on the relatively low number of first- and second-year teachers in these districts in 2006-07, we recruited seven additional districts to join the study in 2007-08. These districts ranged in size from 8,000 to 30,000 students. In one of the Michigan districts, Underwood, only 11 percent of the students were eligible for free and reduced-price lunch and only 11 percent were racial minorities. We decided to include this district in the MIECT study in order to see if there were significant variations in novices’ experiences with mentors and colleagues in a district that was much more affluent than the other.
participating districts. In the other Michigan district, Wagner, 36 percent of the students were eligible for free and reduced-price lunch, making it similar to Whitaker and Kaline in terms of students’ socio-economic status (SES). Table 2 summarizes 2006-07 enrollment and demographic information from the Michigan Department of Education for the two Michigan districts that joined the study in 2007-08.

Table 2. 2006-07 Enrollment and Demographic Information for Two Michigan Districts That Joined MIECT Study in 2007-08

<table>
<thead>
<tr>
<th>District</th>
<th>Total Enrollment</th>
<th>% Free/Reduced Lunch</th>
<th>% Non-white</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underwood (MI)</td>
<td>29,803</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Wagner (MI)</td>
<td>8,242</td>
<td>36%</td>
<td>46%</td>
</tr>
</tbody>
</table>

The five Indiana districts that joined the study in 2007-08 were fairly similar with regard to size and students’ socio-economic status. In terms of size, they ranged from 10,60 students in Payton to 21,715 students in Sayers. With regard to SES, the percentages of students eligible for free and reduced-price lunch ranged from 42 percent in Luckman to 60 percent in Sayers. Finally, the percentages of non-white students in four of the districts ranged from 47 percent to 59 percent, with Payton’s student population including a much higher percentage of non-white students, 83 percent), than the other four districts. Table 3 summarizes 2007-08 enrollment and demographic information from the Indiana Department of Education for the five Indiana districts that joined the study in 2007-08.

Table 3. 2006-07 Enrollment and Demographic Information for Five Indiana Districts That Joined MIECT Study in 2007-08

<table>
<thead>
<tr>
<th>District</th>
<th>Total Enrollment</th>
<th>% Free/Reduced Lunch</th>
<th>% Non-white</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engram (IN)</td>
<td>13,726</td>
<td>59%</td>
<td>47%</td>
</tr>
<tr>
<td>Luckman (IN)</td>
<td>16,153</td>
<td>42%</td>
<td>54%</td>
</tr>
<tr>
<td>Payton (IN)</td>
<td>10,620</td>
<td>47%</td>
<td>83%</td>
</tr>
<tr>
<td>Sayers (IN)</td>
<td>21,715</td>
<td>60%</td>
<td>59%</td>
</tr>
<tr>
<td>Wilson (IN)</td>
<td>12,503</td>
<td>56%</td>
<td>57%</td>
</tr>
</tbody>
</table>

Early Career Teacher Survey Sample. In 2006-07, we invited to participate in the study a) all core content area teachers (math, science, social studies, English/language arts, and elementary general education) and b) all special education teachers in the four Michigan districts (Daus, Greenberg, Kaline, and Whitaker) who taught in grades 1-8 and who were in their first two years of teaching in public schools as certified teachers. In fall 2006 and spring 2007, many of these first- and second-year elementary and middle school teachers completed surveys that asked about their instructional practices; the frequency and substance of interactions with their mentors, colleagues, and administrators; the frequency in which they participated in professional development activities related to reading, writing, and math; and the content of these activities. In addition, these surveys also inquired about the beginning
teachers’ perceptions of social relations within their schools, the manageability of their work, and their future career plans. Further, the spring 2007 survey included measures of novice teachers’ interactions with union representatives, their participation in union activities, and their views on what union priorities should be. Items that asked about teacher background, such as degrees, certification, and university attended, were also included in the spring 2007 survey.

Our survey items on instructional practices in language arts and mathematics are drawn from surveys developed for the Study of Instructional Improvement (SII) at the University of Michigan. Phelps and Schilling (2004) provided evidence of the validity and reliability of the SII survey items that address reading/language arts instruction while Rowan, Harrison, and Hayes (2004) provided evidence of the validity and reliability of the SII survey items that address math instructional practices.

Of the 80 general education teachers in the four districts who were eligible to participate in the study in 2006-07, 58 completed surveys in fall 2006 (n=58; 72.5%) and 32 of these 58 teachers completed surveys again in spring 2007 (n=32; 55.1%). While we were pleased with the response rate in fall 2006, we felt that the response rate in spring 2007 and the low overall number of participants were not fully satisfactory. Thus, as discussed, we recruited seven additional districts (two in Michigan and five in Indiana) to participate in the study in 2007-08, we included first-, second-, and third-year teachers in the sample, and we implemented the Dillman five-contact method in order to attain high response rates (Dillman, 2007). In fall 2007, 255 out of 380 eligible first-, second-, and third-year general education teachers in the 11 participating districts completed fall surveys (n = 255; 67.1%).

Of the 39 special education teachers in Daus, Greenberg, and Kaline who were eligible to participate in 2006-07, 21 completed surveys in fall 2006 (n=21; 53.8%) and 15 of these 21 teachers completed surveys again in the spring (n=16; 71.4%). In 2007-08, we included first-, second-, and third-year special education teachers from nine of the 11 districts in our early career special education teacher sample. In fall 2007, 67 out of 105 eligible first-, second-, and third-year special education teachers in the participating districts completed fall surveys (n = 63.8%).

Mentor/Colleague Sample: In order to collect egocentric social network data, the fall 2006 surveys for beginning general education teachers asked them to list their mentors and other key colleagues with whom they discussed professional issues such as curriculum, instruction, and classroom management. For each novice general education teacher in the sample who completed a fall 2006 survey, we contacted their mentor teacher and up to four of their colleagues (whom they had identified) and invited them to complete a winter 2007 experienced teacher survey. Similarly, the fall 2006 surveys for beginning special education teachers asked them to list their mentors, key colleagues in special education, and key colleagues in general education. For each novice special education teacher who completed a fall
2006 survey, we contacted their mentor teacher, up to four colleagues in special education, and up to four colleagues in general education and invited them to complete a winter 2007 experienced teacher survey.

Therefore, our social network analyses are not based on samples of all of the teachers in participating schools (i.e., the sociometric approach to social network analyses). Instead, our mentor/colleague samples for novice general and special education teachers only include those nominated by the novices as mentors and key colleagues and only those individuals who were responsible for classroom instruction. These mentors and colleagues were surveyed in the spring 2007, and many of the questions asked were the same as those directed to the novice teachers. More specifically, they were asked about their instructional practices, their perceptions of social relations within their schools, the manageability of their work, and their interactions with union representatives, their participation in union activities, and their views of what union priorities should be. Additional background information was also gathered from the experienced teachers, such as years of experience, certification, and degrees.

In collecting data from mentors and colleagues, we focus primarily on individuals who are responsible for classroom instruction for two reasons. First, for a given novice, we are investigating whether the mentor’s and colleagues’ instructional practices are associated with those of the novice (i.e., changes in the novice’s practices over time). If an identified mentor or colleague teaches works in a non-instructional capacity (e.g., school psychologist, guidance counselor), then we would not be able collect data on their instructional practices (and, thus, would not be able to investigate this question). Second, we are investigating how the effectiveness of mentors and colleagues in instruction (as measured by student achievement gains on state tests) is associated with the nature of novices’ instructional practices and their effectiveness (as measured by achievement gains).

**Student Achievement Data.** For each of the 255 early career general education teachers in the 2007-08 survey sample, we are collecting data from the Michigan Department of Education (MDE), the Indiana Department of Education (IDE), and the 11 districts on the performance of their students on state reading, writing, and math tests. For these students, gain scores will be calculated by subtracting pre-achievement test scores from post-achievement test scores (e.g., subtracting 3rd-grade reading scores from 4th-grade reading scores). If it is not possible to create a vertical scale – due to changes in the dimensionality of the tests and possible violations of assumptions of measurement models – we will create residual gain scores (obtained by regressing post-test scores on pre-test scores) and use them as a relative index for ranking teachers in the study. As of 2007-08, MDE and IDE have the necessary data available for calculating student gain scores.

This study features a longitudinal design that employs statistical controls for prior achievement and other student, teacher, and school characteristics (see Ehrenberg et al., 2001; Ferguson, 1998; Rowan, Correnti, & Miller, 2002; Thum, 2003). This design will enable us to test a series of hypotheses about the
effects of mentoring, collaboration with colleagues, and professional development on the outcomes of interest. More specifically, the use of growth models in this design will permit us to use in stages controls for student, teacher, and school characteristics that influence teachers’ instruction and student achievement. For each early career teacher, we will organize the student, teacher, and school information in such a way that the influence of mentoring, collaboration with colleagues or professional development can be tracked. The performance of student cohorts served by this teacher over time will then be structured to test various value-added hypotheses. We will also deal explicitly with the issue of whether, and how adequately, the combination of the data realized and the method of analyses employed in this study support making causal statements (See Frank, 2000).

IV. Plans for Analysis

For this study, we will use multivariate regression and hierarchical linear models (HLM) to study the key outcomes as a function of several independent variables. The main outcomes in the study are early career teachers’ instructional practices, commitment levels, and retention decisions, and the achievement gains of their students in reading, writing, and math. The independent variables include mentor’s instructional practices, colleagues’ instructional practices, nature and frequency of novice’s interactions with mentor, nature and frequency of novice’s interactions with colleagues, expertise of mentor, expertise of colleagues, and content and frequency of professional development. In this section, we first present a model for measuring the effects of these factors on novices’ instructional practices. Then we present a model for measuring change in student achievement as a function of change in novices’ instructional practices. Finally, we describe a strategy for evaluating the robustness of inferences in this study and we briefly discuss issues related to sample size and power.

Effects on Teachers’ Instructional Practices. This study will deploy a series of regression models for estimating the effects of various factors on several instructional practices in reading, writing, and math. That is, we will test a number of hypotheses involving our independent variables and teachers’ practices. One key aspect of instruction, which is a main focus of our study, is the extent to which an early career teacher implements tasks of high cognitive demand (ITHCD). A population model for this aspect of practice is:

\[
(ITHCD) = \beta_0 + \beta_1 \text{ extent to which teacher perceives ITHCD contributes to student achievement} \\
+ \beta_2 \text{ supplemental curricular materials support ITHCD} \\
+ \beta_3 \text{ perceived student ability to engage in ITHCD} \\
+ \beta_4 \text{ professional development supports ITHCD} \\
+ \beta_5 \text{ perceived pressure from mentor, colleagues to emphasize ITHCD relative to other demands}
\]
mentors and colleagues emphasize ITHCD in their own teaching. (5)

The terms in equation (5) are motivated by the reduced form of the utility function in equations (3) and (4). For example, $\beta_1$ through $\beta_4$ represent effects associated with anticipated payoff of effort to develop student human capital, generally representing $\rho$ in the utility function. The parameters $\beta_5$ and $\beta_6$ are examples of effects of conformity to group norms (associated with $1-\rho$ in the utility function), representing the effect of a) perceived expectations to emphasize certain instructional practices and b) mean instructional practices for mentors and colleagues. Note that for each conformity measure, we can also account for the nature and frequency of interactions with each colleague or mentor (see Frank, Zhao, & Borman, 2004; Zhao & Frank, 2003).

While the terms in model (5) represent our central theoretical concerns, other factors could cause teachers to engage in certain practices. Therefore, we will include as controls such factors as mean student performance on state tests, teacher’s university coursework, teacher preparation, teacher’s years of experience, and teacher’s gender and race. Most importantly, after we have obtained longitudinal data, we will control for a teacher’s prior use of ITHCD. Controlling for the prior measure can account for many differences among teachers that are manifest at the time the prior measure is obtained (Allison, 1990). As a result, the terms in (5) represent effects on change in instructional practices.

Of course, we recognize the nesting of teachers within schools and districts. But because our focus is not on comparisons across schools or districts, we will control for between-school differences using fixed effects for those schools in which we have data on more than one new teacher (Wooldridge, 2002); i.e., entering a dummy variable indicator for each school in the study except one. Thus, the model parameters will reflect the effects of intent to develop human capital and conform to social context as experienced by the novices net of any effects that could be attributed to schools or districts.

Effects on Student Achievement in Mathematics. Past research on the use of test scores as indicators of instructional quality has argued against the use of the classroom average score because it represents a “snapshot” that reflects not only students’ instructional histories but also their socio-economic conditions (Bryk & Schneider, 2002; Meyer, 1997). Instead, we propose to proxy the instructional effectiveness of a teacher with measures of how much learning has taken place among her students. Employing state student testing data that are individually associated with teachers, we would estimate the teacher’s instructional effectiveness from the combined performance of the students she has had the opportunity to teach using a series of multilevel growth models (Thum, 2003).

For example, suppose that we use $y_{1j}, y_{2j}, \ldots, y_{ij}, \ldots, y_{r_{ij}}$ to denote the math scores $(t = 1, 2, \ldots, T_{ij})$ of the students $i = 1, 2, \ldots, n_j$ in teacher $j$’s classroom. We would then construct measures
of a student’s achievement status $\pi_{0ij}$ at time $T_j$ and her average rate of learning gains $\pi_{ij}$ using a suitable growth model, such as

$$y_{ij} = \pi_{0ij} + \pi_{ij} \times (t - T_j) + \sum_{i} \delta_i A_{ij} + \epsilon_{ij} \tag{6}$$

for each available testing outcome for each teacher in the sample. These estimates of a student’s growth would also control for the time-varying covariates, such as changes in test form or language program status. These covariates are denoted by $A_{ij}$, and their impact by the corresponding regression weights by $\delta_i$. In most instances, we would assume that these effects do not vary from student to student, hence $\delta_i$ are among the so-called “fixed-effects” that represent constant influence on students.

Since learning growth tends to vary from student to student, a measure of how teacher $j$’s students were achieving would be the precision-weighted average of the measurable amounts of individual learning $i$ over all her $n_j$ students, as represented by the next equation set

$$\pi_{0ij} = \beta_{00j} + \sum_{p>0} \beta_{0pj}(X_{ijp} - \bar{X}_{jp}) + r_{0ij}$$

$$\pi_{ij} = \beta_{10j} + \sum_{p>0} \beta_{1pj}(X_{ijp} - \bar{X}_{jp}) + r_{ij} \tag{7}$$

where an estimate of $\beta_{00j}$ is the average of expected attainments and an estimate of $\beta_{10j}$ gives the average growth rates of her students controlling for student covariates $X_{ijp}$ (such as gender, minority status, etc.). Together, these estimates define for each teacher the pattern of academic performance of her students. Estimates are easily obtained from the solution of the implied mixed-effects model under the assumptions that a) the residual errors $\epsilon_{ij}$ are identically and independently normal with mean 0 and variance $\sigma^2_{ij}$, and b) the parameter residuals $r_{0ij}$ and $r_{ij}$ are distributed bi-variate normal with 0 means, variances $\tau_0$ and $\tau_i$ respectively, and covariance $\tau_{0i}$. The “value-added” estimates, $\beta_{00j}$ and $\beta_{10j}$, describe the combined learning growth experienced by teacher $j$’s students.

After controlling for factors such as aggregate student demographics, Title I status, and English language proficiency, we would then explore how measures of novice teachers’ instructional practices (e.g., the degree to which a teacher emphasizes implements tasks of high cognitive demand (ITHCD)), relate to the academic progress of her students. Note that the measures of instructional practices incorporate the influence of teachers’ professional backgrounds and the effects of their mentors and colleagues. To evaluate this relationship, measures of the instruction (such as ITHCD) experienced by student are posed as predictors in equation 8:
Apart from the teacher-level measures of average student attainment $\beta_{00j}$ and average student growth $\beta_{10j}$, other factors from equation 8 are assumed to be constants. Again, we assumed that the residuals in the teacher-level equation are bi-variate normal with zero means, variances $\psi_0$ and $\psi_1$ respectively, and covariance $\psi_{01}$. 

An interesting issue involving the longitudinal aspect of the study is whether novice teachers change the way they teach (i.e., whether their implementation of tasks with high cognitive demand (ITHCD) evolves), and how this might co-vary with the academic growth patterns of their classes. To explore this, we will employ a modeling procedure similar to Thum (2006) that will relate estimates of the performance of a teacher’s student cohorts to measures of their instructional practices over time.

_Evaluation of the Robustness of the Causal Inferences._ In this study, we will evaluate the sensitivity of our causal inferences in this study to possible violations of the assumptions of inference. Because teachers select themselves into the social contexts we are studying, we will not be certain of the causal inferences we will make about effects of mentors, colleagues, or professional development activities. It may simply be that more effective teachers are able to secure useful assistance, but that there is no effect of these individuals or activities. Our first response to this is to control for a teacher’s prior level of effectiveness. That is, we are estimating the effect of mentors, colleagues, and professional development on change in effectiveness. Furthermore, we also control for important covariates (e.g., changes in students’ backgrounds) that can improve effectiveness.

Aside from these controls, there will be inevitable debate about any causal inferences we make. To respond, we will draw on recent literature to quantify how much the assumptions of inference must be violated to invalidate our inferences (Frank, 2000; Frank & Min, 2007). That is, we will quantify the robustness of our inferences with respect to concerns about omitted confounding variables using Frank’s (2000) impact threshold for a confounding variable. Frank (2000) begins by defining the _impact_ of a confounding variable on an estimated regression coefficient as $r_{v,y} \times r_{v,x}$, where $r_{v,y}$ is the correlation between a covariate, $v$, and the outcome, $y$; and $r_{v,x}$ is the correlation between $v$ and $x$, a predictor of interest (e.g., $x$ is an indicator of mentor effectiveness). Critically, the product $r_{v,y} \times r_{v,x}$ captures both the relationship between the confounding variable and the outcome and between the confounding variable and the treatment.

To obtain the impact of an omitted confounding variable necessary to invalidate an inference, define $r^\#$ as a quantitative threshold for making inferences from a correlation; i.e., $r^\#$ can be defined by a
correlation of a specific magnitude (e.g., an effect size). Here, $r^\#$ is defined by statistical significance. Given the definition of $r^\#$, Frank (2000) suggests that the inference would be invalidated if

$$\text{impact} > \frac{(r_{x\cdot y} - r^\#)/(1-|r^\#|)}{r^\#}.$$  \hspace{1cm} (9)

Thus the quantity $(r_{x\cdot y} - r^\#)/(1-|r^\#|)$ defines the impact threshold for a confounding variable; if there is a confounding variable with impact greater than $(r_{x\cdot y} - r^\#)/(1-|r^\#|)$ then the relationship between the treatment and outcome, given the confound $(r_{x\cdot y|v})$, would fall below the threshold $(r^\#)$ for making a causal inference. Thus the impact threshold will help us quantify the robustness of our inferences to possible misspecification of our models.

Even if an inference is robust with respect to concerns about unmeasured confounding variables, there still may be challenges to the inference based on the generality, or external validity, of the findings. For example, the effect we find in our data may not apply to teachers in other states or at other careers stages. The optimal response would be to randomly sample data from the entire population to which we hope to generalize. But in this study, as in most moderately sized studies, such a sampling scheme was not feasible. This generates a quandary: if we can generalize our results only to the immediate population from which we sampled then our study has limited value for general policy. Does our study have no meaning for those considering effects on teachers in Ohio, or in their fifth year of teaching?

To quantify the robustness of an inference with respect to sample representativeness, Frank and Min (2007) conducted a thought experiment in which part of a sample is replaced with cases from some unobserved sub-population. Defining $\pi$ as the proportion of a sample to be replaced by cases for which the null hypothesis is true, Frank and Min’s calculations then show that the inference is invalid if

$$\pi > 1-r^\#(1-r_{x\cdot y}).$$  \hspace{1cm} (10)

Thus, the quantity $1-1-r^\#(1-r_{x\cdot y})$ defines the index of external validity. Using this index, we can quantify how much of our sample must be replaced by cases for which the null hypothesis is true to invalidate our inference.

Sample Size and Power. As noted, this study will include a large longitudinal database featuring approximately 250-300 novice and 800 veteran teachers. The primary level of analysis will be at the individual teacher level and the sample size will be sufficiently large to enable us to detect moderate effects with power of 0.80 or greater. We will have considerably less power to identify school effects.

V. Significance
This study is significant for a number of reasons. First, it investigates three distinct types of instructional support that shape beginning teachers’ instructional practices, commitment levels and retention decisions, and student learning: a) mentoring, b) collaboration with school-based colleagues, and c) professional development activities. No large-scale studies of induction have followed novices over their first three years of teaching to measure how these factors combine to influence instruction, commitment, retention, and student learning. For example, an early career teacher may participate in Everyday Mathematics professional development, but its long-term effects on her teaching may be influenced by her ongoing interactions with her mentor and colleagues, by the nature of their instructional practices, and by their expectations for her practices.

Second, this study includes both novice general education and novice special education teachers. Therefore, it will enable us to investigate whether the induction experiences of these two groups are similar or quite different. In particular, we will be able to examine the content of both groups’ interactions with mentors and colleagues, and how mentors, colleagues, and professional development seem to affect their commitment levels and retention decisions.

A third reason this study is significant is that we will employ multilevel growth models to investigate how measures of instructional practices are related to student achievement. These measures will incorporate the influence of mentors, colleagues, and professional development on instruction. Such models are ideal for assessing whether early career teachers change the ways they teach and, if so, how such changes co-vary with the achievement patterns of their classes (Thum, 2003).
References


Rowan, B., Correnti, R., & Miller, R.J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *Teachers College Record, 104*(8), 1025-1067.


Endnotes

i Pseudonyms have been used for district names to help ensure confidentiality.

ii As of March 2008, the demographic data for the Michigan districts were not available from the Michigan Department of Education website.

iii With regard to the survey sample of early career special education teachers, we only invited teachers from three of the four districts to participate in 2006-07 (Daus, Greenberg, and Kaline).

iv Due to space limitations, we do not include our full model for measuring the effects of these factors on novices’ commitment levels and retention decisions. For details about this model, see Youngs, Frank, Thum, and Low (under review).

v We are aware that statistical significance is not sufficient for causal inference (Wilkinson et al., 1999). But statistical significance is often the first threshold in a two-step procedure for making causal inferences, “where first the likelihood of an effect (small \( p \) value) is established before discussing how impressive it is” (Wainer & Robinson, 2003, p. 25). I.e., most social scientists are uncomfortable making causal inferences if their estimated effect (or something more extreme) could have occurred more than a small percentage (e.g., 5 percent) of the time by the chance of sampling when in fact the null hypothesis is true.

vi The expressions can be easily adapted to focus on one component correlation when researchers have specific prior beliefs about the strength of the other correlation. The expressions can also be modified to account for the presence of other covariates in the model. See Frank (2000).