Journal of Teacher Education

http://jte.sagepub.com

Intensive Mentoring as a Way to Help Beginning Teachers Develop Balanced Instruction

Randi Nevins Stanulis and Robert E. Floden

Journal of Teacher Education 2009; 60; 112
DOI: 10.1177/0022487108330553

The online version of this article can be found at:
http://jte.sagepub.com/cgi/content/abstract/60/2/112

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
AACTE
American Association of Colleges for Teacher Education (AACTE)

Journal of Teacher Education
Volume 60 Number 2
March/April 2009 112-122

10.1177/0022487108330553
http://jte.sagepub.com
hosted at
http://online.sagepub.com

Intensive Mentoring as a Way to Help Beginning Teachers Develop Balanced Instruction

Randi Nevins Stanulis
Robert E. Floden
Michigan State University

This study examines the impact of intensive mentoring as an induction program component aimed at improving teacher quality in ways that link teaching to student engagement. The Atmosphere, Instruction/Content, Management, and Student Engagement (AIMS) measure of teaching practice, focused on a research-based conception of high-quality teaching known as effective balanced instruction, was used to measure the impact of the intervention. Using a matched comparison group design with 24 beginning teachers, the study tested the effects on teaching practice of intensive mentoring. Findings indicate that the improvement in the beginning teachers' AIMS scores from fall to spring was greater for the experimental group than for the comparison group of teachers.

Keywords: *mentoring; induction; professional development; teacher development*

A support gap exists for novices, especially those who teach in schools of poverty (Johnson, Kardos, Kauffman, Liu, & Donaldson, 2004). Because teacher quality can make an enormous difference in educational opportunities for children (Sanders & Rivers, 1996), beginning teachers need targeted support to overcome the many challenges of learning to teach. In high-poverty districts where turnover is highest, the quality of school cohesion and performance is affected when teachers leave (Ingersoll, 2004). Those responsible for school leadership and policy often do not realize that creating a quality induction program can make a tremendous difference in teacher satisfaction, growth, retention, and impact on students. Only 1% of beginning teachers participate in sustained, comprehensive induction programs (Smith & Ingersoll, 2004). Although the public expects beginning teachers' performance to resemble that of experienced teachers, novices without adequate support need 3 to 7 years of teaching to reach their maximum impact on student learning. Comprehensive induction programs are a way to accelerate this process and minimize the amount of time it takes for a beginning teacher to be most effective in promoting student learning (Alliance for Excellent Education, 2004; Villar, 2004).

Schools face serious challenges as beginning teachers leave before they can develop fully as high-quality teachers. As many as 14% of teachers quit after the 1st year, with numbers rising as high as 50% leaving within 5 years of taking their first teaching position. At the point when a teacher is ready to make a major impact on student achievement, he or she is likely to have left the profession (Alliance for Excellent Education, 2004). Strong induction programs that provide opportunities for teachers to be involved in decision making and that have strong administrative support along with support to develop strong classroom management can keep teachers in the profession (Ingersoll & Smith, 2004). Such programs support organizational and instructional conditions that can help novices develop as high-quality teachers early in their career.

To combat high attrition rates and improve teacher quality, researchers have begun to define the qualities and components of comprehensive induction programs that

Authors' Note: This project was funded by the Carnegie Corporation's Teachers for a New Era and Michigan State University's Literacy Achievement Research Center.

Table 1
Characteristics of Effective Teaching in the Induction Program

Worthwhile Content
Planning deliberate instructional *balance* (direct instruction and group/individual application time), instructional *density* (strong authentic tasks with high expectations, time to discuss and process), and *scaffolding.*
Planning a structure for setting up and maintaining learning tasks and experiences.
Providing opportunities for students to talk, collaborate, and explore content.
Knowing the content beyond the current lesson.
Thinking about content from the students' perspective, anticipating student needs.
Understanding that the quality of the time students spend engaged in their work depends on the quality of the tasks they are expected to accomplish.
Excellent Classroom Management That Engages Students
Creating thoughtfully planned routines during the day, including morning meetings, transitions, routines for group and independent work.
Creating environments that stimulate curiosity; students are absorbed in the work.
Planning for time that is well managed, pacing that is effective for the students' learning. Closure is included in lessons, and much time for interaction, questioning, and discussion is planned.
Paying deliberate attention to both the verbal and the nonverbal environment in the classroom and deliberate attention to both developing relationships with students and among students.
Managing many different kinds of tasks and activities during the day (small group, whole group, one-to-one work with students, discussions, seat and project work).
Strong Motivation and Scaffolding of Student Learning
Developing an environment of high expectations in which students move from dependence to independence in task completion.
Motivating students to spend the time needed to learn complex ideas and solve problems they find interesting.
Anticipating frustrations, segmenting tasks, and providing hints and other mechanisms for students to move from guidance to independence.
Beginning lessons with a "launch" that invites students into a topic and arouses curiosity.

Based on ideas from: Ball (2000); Pressley, Allington, Wharton-McDonald, Block, & Morrow (2001); Stanulis & Manning (2002); Stein, Smith, Henningsen, & Silver (2000); Tomlinson (2003).

make a difference in the lives of teachers and students (Alliance for Excellent Education, 2004; Britton, Paine, Pimm, & Raizen, 2003; Fletcher, Strong, & Villar, 2005; Ingersoll, 2001; Stanulis, 2006). This study examines the impact of a program designed to incorporate several features thought to be important for induction programs aimed at improving teacher quality in ways that link teaching to student engagement. The program focuses on a research-based conception of high-quality teaching known as effective balanced instruction.

Findings from a series of research studies characterized three components of effective balanced instruction specifically in literacy (e.g., Dolezal, Welsh, Pressley, & Vincent, 2003; Pressley et al., 2001; Pressley, Allington, Wharton-McDonald, Block, & Morrow, 2001; Wharton-McDonald, Pressley, & Hampston, 1998), which we have worked to weave with other research and adapt across subject areas (Stanulis & Manning, 2002; Stein, Smith, Henningsen, & Silver, 2000). These three components, teaching of worthwhile content, excellent classroom management that engages students, and strong motivation and scaffolding of student learning, described in Table 1, compose the basis of the vision of teaching that is the intended outcome of the mentoring in this induction program.

Comprehensive Induction Program Components

As a Teachers for a New Era institution (Carnegie Corporation, 2001), we created a specific vision for a comprehensive induction program (Stanulis, 2006). As coprincipal investigators of the Teachers for a New Era project at our institution, the second author (Floden) facilitated cross-college collaborations and a structure to institutionalize this work. As the induction director, the first author (Stanulis) led the design, implementation, and research in the induction program. Part of this induction program involved the evolution of a collaborative partnership between a midwestern urban school district and our university, which was created to pilot-test a comprehensive induction program in a district that anticipated hiring more than 100 teachers for the 2005-2006 school year.

The treatment and comparison groups were similar in some of the induction assistance they received. All beginning teachers in this urban district were invited to

participate in university-led induction experiences. These induction experiences for all new teachers included three half-day orientation sessions held to prepare teachers for this urban context before the school year began, four professional development sessions for the novices during the school year, work with a comprehensive Web-based resource, and a series of seminars for principals. All of the treatment- and comparison-group teachers attended the orientation sessions held at the beginning of the school year. Five teachers from the treatment group attended at least one of the four after-school university-led seminars held across the year; 6 teachers from the comparison group attended at least one of these seminars.

The additional induction support received by teachers in the treatment group was *intensive mentoring*. Those who participated in the treatment group (but not those in the comparison group) interacted weekly with partially released mentors who received intensive preparation from university staff. The purpose of this research study is to compare induction that does not include intensive mentoring to induction that does include a specific model of intensive mentoring based on balanced instruction.

Within the program that includes intensive mentoring, induction is defined as a distinct phase in learning to teach (Feiman-Nemser, 2001a) that "stand[s] as a key juncture of learning, growth, and support" for beginning teachers (Paine, Pimm, Britton, Raizen, & Wilson, 2003, p. 15). At this juncture, the quality of support and learning is critical. The quality of interactions between beginning teachers and their colleagues can play a critical role in the success of novice teachers (Johnson, 2004).

Yet in our experience, we have found that the assignment of mentors is typically based on convenience, volunteerism, and entitlement rather than on selection of mentors who are willing to help novices continue learning to teach (Stanulis, Meloche, & Ames, 2008; Wood & Stanulis, in press). Within the induction program with intensive mentoring, preparation of mentors focused on helping novices enhance student achievement through development of effective balanced instructional practices. Such "educative" mentoring places emphasis on engaging beginning teachers in joint inquiry with a mentor to help novices understand the importance of learning from practice while providing tools useful for studying teaching, including observation, feedback, and analysis of student work (Allen, 1998; Feiman-Nemser, 2001b). This guidance goes beyond sharing instructional tips to place the mentor in a teaching role to build on knowledge and experiences gained at the university. Mentor preparation within this program builds on the research base of Feiman-Nemser (2001a, 2001b), the New Teacher Center (2005; Villar, 2004), and Stanulis et al. (2008), who found that through careful preparation and support, mentors can learn to have instructional conversations with novices that include understanding subject matter (Feiman-Nemser, Schwille, Carver, & Yusko, 1999), planning instruction, student engagement, and formative assessment (Wood, 1999).

Although it is a common component of induction, mentoring is underdeveloped in most school contexts. Our state is typical in having an unfunded state mandate to provide a mentor for each beginning teacher. To comply with this mandate, school districts often assign mentors to provide support for novices without contemplating the complexities of and possibilities for mentoring. Without any preparation, the assigned mentor can become a "buddy," available for advice and explaining school procedures but rarely observing or providing feedback about teaching and learning (Stanulis et al., 2008). Evidence suggests, however, that when mentors have substantial preparation and when the mentoring is instructional and standards based, beginning teachers can also have a significant impact on student achievement (Fletcher et al., 2005).

The intensive mentoring component involved 5 teachers released 1 day each week to mentor 3 beginning teachers. Current classroom teachers were recruited and interviewed for mentor positions. Mentor–novice matches were based on teaching responsibilities related to content and level—secondary science, English, and math; special education; and elementary education. Developing mentoring as a practice required intensive, sustained preparation; therefore, mentors participated in mentor study groups for 6 hr each month along with 6 full days of professional development during the school year. In addition to meeting weekly with their assigned beginning teacher in the novice's classroom to observe and provide feedback, coplan, analyze student work, or demonstrate a teaching practice, the mentors also led a monthly seminar with their group of novices. These study groups provided a time for beginning teachers to connect with one another and for the mentor to facilitate discussions about continued learning to teach. Finally, university-based coaches support the school-based mentors by observing the mentors in their conversations with beginning teachers (Stanulis, 2006).

Developing as Effective Teachers

Mentor preparation focused on helping the mentors work closely with beginning teachers to support their development as effective teachers of balanced instruction. Within mentor study group sessions, there was a deliberate effort to have university and school personnel share a vision of effective teaching that emphasized developing classroom practices that balance holistic learning activity and skill instruction, creating effective management that

engages learners, and designing instruction that promotes student academic motivation through reflective analysis of planning, instruction, and assessment (Roehrig, Dolezal, Mohan, Bohn, & Pressley, 2003). Mentors learned ways to work with novices in managing a class, planning and implementing engaging learning tasks, knowing subject matter, assessing student learning, and learning in and from their practice as teachers.

Methods

The strongest design for testing the effectiveness of this program would be one where beginning teachers were randomly divided into two groups—an experimental group, with both intensive mentoring and the regular district induction, and a comparison group, with only the regular district induction. But within the realities of working in this particular school district, where union and district administrators required that the experimental intensive mentoring be done only with volunteers, we were not able to randomly assign new teachers to treatment and comparison groups. To address the concern that selection bias occurred because volunteer participants were different from beginning teachers who received only the regular district induction, we selected teachers for our comparison sample by choosing 1st- or 2nd-year teachers who matched treatment participants on two relevant characteristics: subject area and years of experience working in the district. We were able to select comparison teachers who matched on both these dimensions with one exception, a treatment-group English teacher for whom we could find a match in subject area but not in years of experience. All of the participants in both groups graduated from a traditional teacher preparation program. There were not enough males in the sample size to match along gender (i.e., there was only 1 male elementary teacher and there were only 2 male secondary teachers). We also examined whether the two groups were teaching students from similar backgrounds; as described in more detail later, the two groups of teachers taught in schools with similar proportions of students from minority groups and students on free and reduced lunch.

Research Design

We selected a measure of teaching practice matched to our program's substantive focus on balanced instruction. Our research hypothesis was that beginning teachers in the program with intensive mentoring would make greater gains in their teaching effectiveness than beginning teachers who had only the regular district induction program. With this matched-pairs design, we observed and rated 24 beginning teachers' practice (12 treatment, 12 comparison) at both the beginning and the end of a school year. For the observation, we used the Atmosphere, Instruction/Content, Management, and Student Engagement (AIMS) observation instrument, which produced ratings along four dimensions of classroom outcomes—atmosphere, instruction and content, management, and student engagement (Roehrig, 2003). Scales for each dimension are based on multiple observation items, each rated on a 3-point scale.

An illustrative item for the Atmosphere subscale is "Teacher encourages curiosity/suspense—getting students excited about what they are learning/doing." A sample item for Instruction/Content is "Teacher provides many opportunistic mini-lessons at teachable moments." For Management, one item was "Teachers gives explanation/rationale for decisions/rules/requests." For Student Engagement, one item was "At least 80% of students are consistently on task and highly engaged in class activities."

Participants

Members of the treatment group were 12 first- and second-year teachers in a single midsized, economically depressed, urban district. This group of volunteers included 3 teachers in English (secondary), 3 in mathematics (secondary), 3 in elementary education, 2 in science (secondary), and 1 in special education, with each of the subgroups assigned to an intensive mentor within the subject matter or area of expertise. Twelve other beginning teachers participated in the project as the comparison group, receiving the regular district induction interventions (orientation, after-school seminars, principal seminars) but no "intensive mentor." The comparison group received the typical "building mentor" as assigned by the individual school principal. The mentors in the comparison group were not matched by subject matter. These mentors received no specific preparation for their role and were not expected to observe in the beginning teacher's classroom or provide feedback on their developing practice. Mentors in the comparison group schools were not released to spend time with the beginning teacher during the school day. The 12 beginning teachers in the comparison group, each matched to a teacher in the experimental group, agreed to be observed on the same schedule as the experimental group.

Participants for the treatment and comparison groups were all beginning teachers entering their 1st or 2nd year of teaching. All attended the university-led orientation sessions held before the school year began. Nine of the matched pairs were 1st-year teachers. Two of the secondary mathematics matched pairs were 2nd year teachers. One of the English matched pairs included a 1st-year

teacher (treatment) and a 2nd-year teacher (comparison group). During the orientation session before the school year began, we held two different informational meetings in small groups with the beginning teachers, outlining the program participation option of participating in either the comparison or the treatment group. The comparison group consented to be observed at the beginning and end of the school year. The treatment group consented to the same observation, with the addition of weekly intensive mentoring in their classroom. Of the 102 teachers the district hired, only 43 true beginning teachers (i.e., in their 1st or 2nd year of teaching) attended our orientation session, thus restricting our pool. The district does not espouse a specific vision of teaching, but a pacing guide regulates the kind of practice and pace that teachers should follow. The professional development that we provided focused on becoming an effective urban teacher, working within the context of their curriculum, while introducing new ideas and practices. Because this program was not a requirement of the district and was promoted mostly by the university faculty, the participants had no district incentive to participate.

Some beginning teachers were apprehensive about having a mentor visit their classroom regularly, stating that they were finally "free from this regular scrutiny" since graduating from their teacher preparation programs. Others had already found a building mentor with whom they had bonded and were reticent about entering a new relationship with another mentor. We began the study with 15 matched pairs, but because of changes in teaching assignments and comparison-group members dropping out of the study, we collected complete data from 12 matched pairs.

The principals from schools where beginning teachers had been hired were invited to three principal breakfast seminars led by university faculty. These seminars focused on learning about ways to help beginning teachers thrive, ways to provide substantive feedback to beginning teachers, and ways to think about fostering educative mentoring in their buildings. Principals of 8 of the comparison-group teachers and of 10 of the experimental-group teachers attended the seminars. (At the secondary level, principals had both comparison- and treatment-group teachers in their buildings.)

According to their logs, the intensive mentors observed each beginning teacher 29 to 31 times across the school year. During these sessions, the mentors coplanned, observed and provided feedback, taught demonstration lessons, went together to observe another teacher, analyzed student work, and designed curriculum maps. Mentors also helped the beginning teachers garner resources needed to teach (textbooks) and survive within difficult working conditions (e.g., being the only classroom in the school without air conditioning, having the most challenging group of first graders, teaching high school physics when the teacher was certified to teach English, moving between two buildings on a lunch hour, teaching on a cart without a classroom, teaching in a context that does not seem to value a teacher's ideas, handling midyear displacement to another building and grade level). Each mentor held eight beginning-teacher study groups after school during the year. Each intensive mentor and beginning teacher in the treatment group also attended a daylong retreat where together we studied and discussed issues of teaching in an urban school.

The comparison- and treatment-group participants taught in very similar urban contexts. In the high schools where teachers from both groups taught, free and reduced lunch ranged from 50% to 57% and the percentage of minority students ranged from 57% to 79% of the school population. At the middle school where teachers from both groups taught, free and reduced lunch ranged from 65% to 75%, and percentage of minority students ranged from 62% to 79%. At the elementary level, free and reduced lunch ranged from 51% (where a comparison teacher taught) to 91% (where a treatment teacher taught). The percentage of minority students ranged from 50% to 81%.

The AIMS Instrument

Each teacher was evaluated using the AIMS tool, developed by a team of literacy researchers led by Alysia Roehrig (Roehrig et al., 2003). This tool measured 130 criteria gathered by observing teaching practice and student behavior in the classroom. Specifically, Roehrig and her colleagues (2003) reported that a high score on the AIMS instrument was matched with effective teaching practices, including creating rich learning tasks, atmospheres where learning was valued, and excitement about academic content and activities, and having few discipline problems because students were on task. The AIMS instrument has been used in research that examined effective and less effective primary grades teachers (Bohn, Roehrig, & Pressley, 2004), mentoring beginning primary teachers for exemplary teaching practices (Roehrig, Bohn, Turner, & Pressley, 2008), and associations between classroom practice and preservice teachers' knowledge about effective primary reading instruction (Roehrig, Guidry, et al., 2008).

The 130 items on the AIMS instrument were based on qualitative research on exemplary teachers, including observation of classroom teachers and surveys completed by exemplary teachers and university researchers (Roehrig et al., 2003). The AIMS items were grounded in analysis

of classroom practices of teachers who were successful in maintaining high levels of student engagement and instructional improvement gains in diverse students (e.g., Allington & Johnston, 2002; Bogner, Raphael, & Pressley, 2002; Day, Boothroyd, Johnston, & Cedeno, 1999; Pressley et al., 2001). Although originally used to identify literacy practices of exemplary teachers, these practices can be described and measured in ways appropriate for a range of subject areas. They deal with broad categories, including classroom atmosphere, classroom instruction, classroom management, and student engagement. Findings from this research indicated that the more teachers integrated practices similar to those observed in exemplary classrooms, the more students were highly engaged in classroom learning (Roehrig et al., 2003).

The reliability of the instrument was measured by having multiple raters observe 16 different elementary teachers with variable teaching experience, with each of the 2 to 3 raters observing 3 to 7 hr of instruction. Independently, raters indicated whether they observed a practice *seldom* (1), *somewhat* (2), or *consistently* (3). For individual items, the interrater agreement was 65%.

Roehrig and her colleagues (2003) conducted studies to assess multiple aspects of the validity of inferences from scores on the AIMS instrument to conclusions about teaching quality. Examination of items and scales by academic experts and effective elementary teachers established face validity. Evidence of construct validity was provided by computing Cronbach alpha statistics for each of the four scales. The alpha values for the scales ranged from .60 to .96, showing that the scales were internally consistent. Construct validity was also examined by computing correlations between student engagement (considered as an outcome variable) and the other three scales (considered as measures of teaching quality). The correlations were all positive and statistically significant.

The validity of the instrument was further tested for this study. Because the AIMS instrument was initially designed for observation of elementary classrooms, a second research associate accompanied the primary research associate to study a secondary English teacher and a secondary science teacher across several class periods to confirm the validity of the instrument for secondary school classrooms and identify items not appropriate for observations of secondary instruction. The two raters independently identified the same seven items to be deleted when observing secondary teachers. These items, judged to be inapplicable to a secondary school context, were on features of instruction such as art tie-in (e.g., the art connected with the lesson), direct phonics instruction (e.g., teacher embeds direct phonics instruction, and teacher points out and has students find the chunks in words), play embedded in instruction, and stretching words (e.g., teacher models stretching words, teacher encourages students to stretch words). To confirm that the coding of the other items remained reliable, two observers used the instrument in two different secondary classrooms, with an interrater agreement of 85%.

Data Collection

To measure the effectiveness of the intervention, a researcher, blind to the condition of each of the 24 teachers, observed all of the teachers (both the experimental and the comparison groups) in September to October and April to May of the school year. In preparation for data collection, a second research associate accompanied the primary research associate to pilot score two beginning elementary teachers and two beginning secondary teachers who were not part of this study. The two researcher associates then discussed their scoring differences and ways they would distinguish whether a beginning teacher's instruction would fit under *seldom*, *somewhat*, or *consistently* in each of the 130 items on the instrument. After each observation, the two raters discussed items where there was disagreement. In each case, the raters were able to reach agreement.

Survey Instrument

At the end of the academic year, the beginning teachers in the treatment group were asked to complete an online survey. Although we asked the participants to identify themselves, we told them that their individual responses would not be reported to their mentor and that their responses would be reported only via pseudonyms. All 12 beginning teachers responded to the open-ended survey questions:

- What specific activities were you involved in with your mentor during your individual meetings (e.g., observations and feedback, coplanning)?
- What were the benefits and disadvantages of the monthly study groups (held by your mentor teacher) to your learning?
- What specifically did you learn this year from your mentor that directly helped your teaching practice?
- Describe the benefits (if any) of participating in the induction program.
- Describe what you wish you could have learned from your mentor that you did not learn.
- Describe whether you were able to have open and candid conversations with your mentor and why (or why not).

Table 2
Pretest Means and Standard Deviations for Total Atmosphere, Instruction/Content, Management, and Student Engagement (AIMS) Scale and Four Subscales

	Experimental Group		Comparison Group	
Scale	*M*	*SD*	*M*	*SD*
AIMS total	1.9725	0.1562	2.0678	0.1951
Atmosphere	1.9924	0.1711	2.1155	0.2011
Instruction/Content	1.9901	0.1145	2.0594	0.1926
Management	1.9149	0.3840	2.0192	0.3298
Student Engagement	1.6042	0.5687	1.8056	0.3064

Table 3
Posttest Means and Standard Deviations for Total Atmosphere, Instruction/Content, Management, and Student Engagement (AIMS) Scale and Four Subscales

	Experimental Group		Comparison Group	
Scale	*M*	*SD*	*M*	*SD*
AIMS total	2.1153	0.1601	2.0371	0.1880
Atmosphere	2.1155	0.1719	2.0704	0.1822
Instruction/Content	2.1166	0.1308	2.0327	0.1730
Management	2.1142	0.4239	1.9964	0.3086
Student Engagement	2.0208	0.4325	1.8542	0.5163

Because the survey questions involved data about specific interactions with the intensive mentors within this induction program, beginning teachers who participated in the comparison group were not asked to complete this survey.

Quantitative Analysis and Results

The two questions we addressed were (a) Were the AIMS scores for treatment and comparison groups equal in the fall, before intensive mentoring began? and (b) Was improvement in the beginning teachers' AIMS scores from fall to spring greater for the experimental group than for the comparison group?

Descriptive statistics of the teachers' scores on the AIMS tool for the comparison group and experimental are displayed in Table 2. The teachers' average score on the AIMS tool in the fall for the comparison group (M = 2.07, SD = .20, n = 12) was slightly higher than for the experimental group (M = 1.97, SD = .16, n = 12). Our concern about selection bias, that better teachers might have volunteered to participate in our program, was negated because the comparison group began the school year with higher scores on the AIMS instrument. That initial superiority of the comparison group was also found for each of the four subscales.

In spring, teachers in the experimental group (M = 2.11, SD = .16, n = 12) demonstrated higher average scores on the AIMS tool compared to those in the comparison group (M = 2.04, SD = .19, n = 12). That difference was also found for each of the subscales (see Table 3). To answer our second question, we wanted to look at differences in gain between the experimental and the comparison groups. Given the modest sample sizes, we were concerned that the distribution of scores might not fit a normal distribution, so we used the nonparametric Wilcoxon sign test to test for statistical significance for difference between the experimental group and the comparison group in their fall-to-spring gains. The Wilcoxon test uses a statistic based on rank ordering of matched pairs of observations (in this case, rank ordering them on the difference in fall-to-spring gain). The Wilcoxon statistic is approximately normally distributed for eight or more matched pairs. We conducted the test for both the AIMS instrument as a whole and for each of the four subscales.

Table 4
Wilcoxon Signed Ranks Test for Total Atmosphere, Instruction/Content, Management, and Student Engagement (AIMS) Scale and Four Subscales

Scale	*N*	Number of Negative Ranks	Sum of Negative Ranks	*Z*
AIMS total	12	2	6.0	–2.589*
Atmosphere	12	2	8.5	–2.394*
Instruction/Content	12	1	3.5	–2.786**
Management	12	3	19.5	–1.201
Student Engagement	12	3	7.5	–2.055*

*p < .05. **p < .01. ***p < .001.

Table 4 presents the results of the Wilcoxon tests. Using a two-tailed test of significance (p = .05), differences favoring the experimental group were found for the overall AIMS score and for three of the four subscales. The one subscale for which the difference was not statistically significant (although still favoring the experimental group) was Management.

Qualitative Analysis and Results

The procedures for analysis of the 12 surveys involved the following process. First, we read all of the comments

for a given question. Next, we completed a holistic summary of the comments for each question. As we developed the summary for a question, we looked for overall patterns as well as frequency of responses. Finally, we selected direct quotes that illustrated the general tenor of all of the comments. The analysis provides comments that were representative of the general ideas provided in the responses from the 12 beginning teachers. These comments include both the comments reported most frequently as well as the range of the comments provided.

In the surveys, the beginning teachers described the mentoring activities they were involved in during the year (all names are reported as pseudonyms). The mentoring activities that were mentioned most frequently included coplanning, observation and feedback, and providing management strategies. An elementary teacher, Mary, wrote that her mentor "provided thoughtful feedback on lessons, analyzed student work and behavior, and used a cycle of preplanning, observations, and then reflections about lessons observed." John, an English teacher, wrote about how he and his mentor "worked on lesson planning and classroom management, and ways to locate instructional activities for differentiated learning styles." Andrea, a science teacher, wrote, "We met to discuss student behavior and analysis of student work. We had coplanning sessions that were very helpful."

There were several benefits and disadvantages mentioned regarding participation in after-school seminars held with each group of beginning teachers and their mentor. The benefits that were mentioned most frequently included an opportunity to share ideas, resources, and advice; an opportunity to hear from other new teachers who were going through similar struggles; and the increased openness to try new things in their practice that resulted from discussions in the study groups. The disadvantages that were mentioned most frequently involved feedback that the seminars came at a time after school when the beginning teachers were feeling exhausted, stressed, and pulled with many other demands. Through participation in the eight after-school beginning teacher seminars, mathematics teacher Brian stated, "I was challenged to think about increasing my openness to trying things in my classroom that I otherwise may not have thought about trying, such as having discussions about racial/ethnic diversity and integrating other subjects into my teaching." Barb, an elementary teacher, found the study groups "a place for great discussions between beginning teachers and a veteran teacher . . . [where we] discussed topics such as differentiated instruction, guided reading, and diversity in the classroom. We learned a lot from each other's knowledge." Janet, a science teacher, found that through the study groups, "the new ideas we talk about get me to do some research into the topics to enhance my teaching." Kathy, an English teacher, said that the study group meetings "pushed me and my thinking about teaching, and they were simply great motivators to keep going." Marilyn, a middle school special education teacher said, "Sometimes after a long day, the meetings came at just the time you felt burnt out and ready to go home." She did find that "we were able to share articles and resources to expand my thinking about teaching in an urban area."

During this 1st year of the mentoring program, some of the beginning teachers did not feel that enough time was spent talking about content knowledge. Brian, a mathematics teacher, wrote, "Not much time was spent discussing content knowledge," and Sarah and Janet, both science teachers, said that there was "not really anything" that they specifically learned about content. Mary (elementary), Edna (mathematics), Barb (elementary), Erica (mathematics), and Kathy (English) wrote that they felt that their work with their mentor did enhance their content knowledge.

Although some of the novices could not pinpoint specific conversations about "content," they did write about ideas and stances that they learned from their mentors that directly helped improve their teaching. In particular, the beginning teachers wrote about ways in which their mentor pushed their thinking, helped them plan, and helped them engage students. John (English) described how his mentor "provided materials about varied instructional methods to address differentiated learning styles to increase student engagement." Ann (science) wrote, "My mentor observed my room and kept a detailed log of student engagement. I got to find out many things that I didn't notice in my room." Mary (special education) wrote about how "she encouraged innovative and engaging ideas and activities [and] emphasized when students were not engaged." Mary (elementary) described how her mentor "pushed my thinking as to why I make the choice I do when teaching," and Sarah (science) wrote, "We would review the goals of the lesson, and I realized that I did not actually meet those goals of that same lesson." Kathy (English) wrote, "I learned about the need to vary instruction to suit different reading levels and attention spans."

Many of the beginning teachers wrote about ways that they worked with their mentors to analyze student work and to collect data about student learning. Mary (elementary) wrote,

> My mentor suggested I create a system to keep all the data I collect for a particular student. She helped me evaluate running records and we had discussions on how to use the information I observed to inform my instruction. We also

> had a discussion about "authenticity" with assessments and how much value to place on just one type of assessment.

Andrea (science) wrote, "We discussed differentiated assessment so that all types of learners could succeed. As a result, I incorporated projects, papers, test [*sic*] and papers into my assessment curriculum." Janet (science) wrote, "I learned that I need to set higher expectations for the students and hold them accountable."

When asked to describe the greatest benefits of having an intensive mentor throughout the school year, the beginning teachers wrote about many ways that they had grown. All of the beginning teachers agreed to participate in the intensive mentoring for a 2nd year. None of the beginning teachers felt that they had not benefited from this joint work. Mary (elementary) wrote, "I would have felt very alone without a mentor teacher this year. . . . It was very helpful to have an open-minded, unbiased, experienced observer provide feedback and advice. Debbie was my saving grace this past year." Similarly, Barb (elementary) wrote, "There is nothing I can say my mentor hasn't done for me. . . . She has supported me through a very trying year. The first week of school I wanted to quit. She made me believe in myself and want to continue to grow as a teacher." Brian (mathematics) wrote,

> Even though I was not always excited to discuss the seemingly many issues I was having in my classroom, in retrospect I am glad that I had the opportunity to discuss those areas needing improvement. I'm sure my teaching was much better than it would have been without this program.

Kathy (English) wrote, "Wow! Having support from Peggy [mentor] and others involved in the program helped me so much in realizing how much I could improve my practice."

The beginning teachers also cited ways in which they felt the mentoring could improve. When asked what they wish they could have learned from their mentor that they did not, the beginning teachers most frequently said that they would have liked the opportunity to observe their mentor teach. Brian (mathematics) wrote, "I would have liked to watch her teach on more than one occasion." Sara (science) wrote, "I wish she could have pointed out more things for me to work on." Mary (special education) wrote, "I would like more opportunities to look at student work together and assess it." John (English) wrote, "The information and materials that we received were excellent and will be of great value next year in developing my teaching. It was hard to impossible to assimilate the material while struggling with the first year of teaching."

Discussion

The qualitative reports of the treatment group indicate that the induction program provided multiple opportunities to learn. In addition, the intensive-mentoring induction intervention had an effect on gains in teacher effectiveness as measured by the AIMS tool. Although the comparison group began the school year with higher scores, the experimental group rose above comparison by the end of the school year, after 8 months of intensive mentoring and participation in induction seminars with university personnel. On three of the four subscales (Atmosphere, Instruction/Content, and Student Engagement), a two-tailed test of significance favored the experimental group. Results for the Management subscale favored the experimental group, although the differences were not statistically significant.

Categories of Effective Teaching

This study shows that intensive mentoring focused on balanced instruction improved teaching practice, as measured by an observation tool aligned with the specific program goals. The categories of effective teaching that composed the intended intensive mentoring curriculum included teaching worthwhile content, classroom management that engages learners, and strong motivation and scaffolding of student learning. It was our hope that deliberate study of these dimensions of teaching practice through seminars and intensive mentoring would help beginning teachers' practice improve along these dimensions. This intensive mentoring involved close work in the classroom, where mentors observed, coplanned, analyzed student work, and collected and analyzed teaching data together with a beginning teacher.

From analysis of the data, we wonder why the AIMS management scores in the experimental group did not rise significantly even though management was one focus of the intensive mentoring. One possible explanation is tied to the fact that the mentoring component of our program did not begin until 3 weeks into the school year because of hiring and placement issues in the district. As other research shows, initial management routines established in the first weeks are difficult to change (Evertson & Smithey, 2000). It is critical for an induction program to begin the mentoring component before or during the 1st week of school.

Another possible explanation for the AIMS management scores could be the particular focus of the induction program. In typical mentoring, management is a dominant focus of the work with beginning teachers. In the induction treatment, mentors focused on helping

beginning teachers develop instructional practices that extend beyond managing behavior in classrooms. Because the curriculum of the induction program emphasized balanced instruction that integrated engagement with students through strong content, some critical elements of management (routines, procedures, timing, consequences) could have been emphasized less in the treatment group.

Limitations of the Study

The experimental and comparison groups taught in comparable teaching contexts, serving similar student populations within the same urban district. Although we acknowledge that teachers who volunteer to participate in this kind of treatment are more likely to benefit than those who might be required to participate, we do not have any evidence that the experimental group of teachers needed more help than the comparison teachers. Still, we acknowledge that it is possible that the teachers who volunteered to participate may have been more open to feedback, limiting the generalizability of the gains they made with intensive mentoring.

The fiscal situation in this district created challenges to studying the long-term effects on teacher quality and retention. At the beginning of the year of our study, the district hired more than 100 beginning teachers, had an unprecedented midyear layoff, and then laid off 65 of these teachers at the end of the school year, including most of the beginning teachers in our treatment and comparison groups. The intent of the original research was to study these novices across their first years of teaching in this district to measure the year-to-year gains in teacher effectiveness. The teacher layoffs made it impossible to continue our comparisons between these two groups of beginners.

We were able to maintain contact with the teachers in the treatment group. Remarkably, all of these teachers remained in teaching for the next school year, whether in or outside of the district. That rate of retention dramatically exceeds national averages. We were not, unfortunately, able to determine how many of the beginners in the comparison group also remained in teaching.

Although the preliminary results are positive, there is more to be learned about both teacher quality and teacher retention. We wonder if the isolation of learning to teach and possible lack of support could have hindered promising teachers in the comparison group from making more improvement. More data about the experiences of comparison group teachers need to be collected in future studies. In addition, we realize that data linking teacher effectiveness to student performance are needed. This study takes a first important step in defining and analyzing teacher effectiveness.

Next Steps

Next steps in trying to understand the effects of intensive mentoring include attempts to conduct a study with random assignment by locating a district willing to randomly assign beginning teachers to either the treatment or the comparison group. Another way to conduct a larger scale random assignment would be through cross-district comparisons, in which some districts receive the intensive mentoring treatment and other districts maintain their current induction activities. (When assignment of treatment is done at the district level, however, getting enough districts to secure adequate statistical power for cross-group comparisons is difficult.) Student pre- and posttreatment tasks aligned with the characteristics of effective teaching could provide ways for us to trace a path from teacher effectiveness to student achievement.

As a result of this initial study, we now have some evidence that with intensive induction, beginning teachers can improve in areas of teacher effectiveness. With a larger sample and after more time, more can be learned about ways in which teacher effectiveness can be accelerated with the support of a strong, comprehensive induction program. Because a major challenge is designing induction support that applies across the range of subject areas, it is desirable to establish programs with a research-based focus that can be used across subject areas, with matching along subject areas for program components as well.

References

Allen, D. (1998). *Assessing student learning: From grading to understanding.* New York: Teachers College Press.

Alliance for Excellent Education. (2004). *Tapping the potential: Retaining and developing high-quality new teachers.* Washington, DC: Alliance for Excellent Education.

Allington, D. L., & Johnston, P. H. (2002). *Reading to learn: Lessons from exemplary fourth-grade classrooms.* New York: Guilford.

Ball, D. (2000). Forward. In M. K. Stein, M. S. Smith, M. A. Henningsen & E. Silver (Eds.), pp. ix-xiv. *Implementing standards-based mathematics instruction* (pp.). New York: Teachers College Press.

Bogner, K., Raphael, L. M., & Pressley, M. (2002). How Grade 1 teachers motivate literate activity by their students. *Scientific Studies of Reading*, *6*, 135-165.

Bohn, C., Roehrig, A., & Pressley, M. (2004). The first days of school in the classrooms of two more effective and four less effective primary-grades teachers. *Elementary School Journal*, *104*(4), 269-287.

Britton, E., Paine, L., Pimm, D., & Raizen, S. (2003). *Comprehensive teacher induction: Systems for early career learning.* Norwell, MA: Kluwer Academic.

Carnegie Corporation of New York. (2001). Teachers for a new era. Retrieved December 28, 2006, from http://www.carnegie.org/reporter/03/foundation_roundup/newera.html

Day, J. P., Boothroyd, K., Johnston, P., & Cedeno, M. (1999, December). *Larger fourth grade study findings in exemplary teaching practices*. Paper presented at the National Reading Conference, Austin, TX.

Dolezal, S. E., Welsh, L. M., Pressley, M., & Vincent, M. M. (2003). How nine third-grade teachers motivate student academic engagement. *Elementary School Journal*, *103*, 239-267.

Evertson, C. M., & Smithey, M. W. (2000). Mentoring effects on proteges' classroom practice: An experimental field study. *Journal of Educational Research*, *93*, 294-304.

Feiman-Nemser, S. (2001a). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, *103*(6), 1013-1055.

Feiman-Nemser, S. (2001b). Helping novices learn to teach: Lessons from an exemplary support teacher. *Journal of Teacher Education*, *52*(1), 17-30.

Feiman-Nemser, S., Schwille, S., Carver, C., & Yusko, B. (1999). *A conceptual review of literature on new teacher induction*. National Partnership for Excellence and Accountability in Teaching. Unpublished manuscript, East Lansing, MI.

Fletcher, S., Strong, M., & Villar, A. (2005). *An investigation of the effects of variations in mentor-based induction on the performance of students in California*. Paper presented at the seventh national New Teacher Center Symposium, San Jose, CA.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499-534.

Ingersoll, R. M. (2004). *Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers?* Washington, DC: Renewing Our Schools, Securing Our Future, A National Task Force on Public Education.

Ingersoll, R. M., & Smith, T. (2004). Do teacher induction and mentoring matter? *NASSP Bulletin*, *88*(638), 28-40.

Johnson, S. M. (2004). *Finders keepers: Helping new teachers survive and thrive in our schools.* San Francisco: Jossey-Bass.

Johnson, S. M., Kardos, S. M., Kauffman, D., Liu, E., & Donaldson, M. L. (2004). The support gap: New teachers' early experiences in high-income and low-income schools. *Education Policy Analysis Archives*, *12*(61). Available from http://eepa.asu.edu

New Teacher Center. (January, 2005). *Mentor Academy handouts*. Presentation to Michigan State University, East Lansing.

Paine, L., Pimm, D., Britton, E., Raizen, S., & Wilson, S. (2003). Re-thinking induction: Examples from other countries. In M. Scherer (Ed.), *Keeping good teachers* (pp. 67-80). Alexandria, VA: ASCD.

Pressley, M., Allington, R., Wharton-McDonald, R., Block, C., & Morrow, L. (2001). Learning to read: Lessons from exemplary first-grade classrooms. New York: Guilford.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., et al. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, *5*(1), 35-58.

Roehrig, A. D. (2003). *The effects of mentoring on beginning teacher effectiveness and student outcomes*. Doctoral dissertation, University of Notre Dame, Notre Dame, IN.

Roehrig, A. D., Bohn, C. M., Turner, J., & Pressley, M. (2008). Mentoring beginning primary teachers for exemplary teaching practices. *Teaching and Teacher Education*, *24*(3), 684-702.

Roehrig, A. D., Dolezal, S. E., Mohan-Welsh, L., Bohn, C. M., & Pressley, M. (2003, April). *The development of a tool to assess the quality of early-primary grade teachers' classroom practices*. Poster presented at the annual meeting of the American Educational Research Association, Chicago.

Roehrig, A. D., Guidry, L., Bodur, Y., Guan, Q., Guo, Y., & Pop, M. (2008). *Guided field observations: Variables related to preservice teachers' knowledge about effective primary reading instruction*. Unpublished manuscript.

Sanders, W., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future students' academic achievement. Knoxville: University of Tennessee, Value Added Research and Assessment Center.

Smith, T., & Ingersoll, R. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, *41*(3), 681-714.

Stanulis, R. N. (2006, April). *Keeping content and context central: Comprehensive induction in the Michigan State University Teachers for a New Era project*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Stanulis, R. N., & Manning, B. (2002).The teacher's role in creating a positive verbal and nonverbal environment. *Early Childhood Education Journal*, *30*(1), 3-8.

Stanulis, R. N., Meloche, B. H., & Ames, K. T. (2007, April). *Challenges in developing, enacting, and analyzing mentoring that focuses on teaching and learning: A university–school collaboration*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. New York: Teachers College Press.

Tomlinson, C. A. (2003). *Fulfilling the promise of the differentiated classroom: Strategies and tools for responsive classrooms*. Alexandria, VA: ASCD.

Villar, A. (2004). *Measuring the benefits and costs of mentor-based induction: A value-added assessment of new teacher effectiveness linked to student achievement*. Santa Cruz, CA: New Teacher Center.

Wharton-McDonald, R., Pressley, M., & Hampston, J. M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal*, *99*, 101-128.

Wood, A. L. (1999). How can new teachers become the B.E.S.T.? In M. Scherer (Ed.), *A better beginning: Supporting and mentoring new teachers* (pp. 116-123). Alexandria, VA: ASCD.

Wood, A. L., & Stanulis, R. N. (in press). Quality induction: Fourth wave induction programs. *New Educator Journal*.

Randi Nevins Stanulis is an Associate Professor at Michigan State University and is the faculty leader of the MSU Teacher Induction Program. She received Michigan State University's Outreach Scholarship Community Partnership Award in recognition of MSU's induction collaboration with an urban district. She has studied mentoring in learning to teach, beginning teacher learning, and learning to teach literacy.

Robert E. Floden is a University Distinguished Professor at Michigan State University and Co-PI of MSU's Teachers for a New Era initiative. He has studied teacher education and other influences on teaching and learning. His current research examines teacher preparation and teachers' mathematical knowledge for teaching.