Informing Teacher Education through the Use of Multiple Classroom Observation Instruments

By Nancy Dubinski Weber, Hersh C. Waxman, Danielle B. Brown, & Larry J. Kelly

One of the primary goals of the No Child Left Behind Act of 2001 (NCLB) is increasing student achievement by holding schools, districts, and states accountable for academic growth. Under threat of governmental intervention, schools must reach adequate yearly progress, which measures annual standardized test scores and graduation rates to assess how the overall student population, as well as key demographic student groups, performs regarding state academic content standards. In 2009, the Race to the Top legislation placed further pressure on educators to raise achievement as it called for the use of data-driven instructional practices and mandated

Nancy Dubinski Weber is a research associate in the Education Research Center and a lecturer in the Department of Teaching, Learning, and Culture of the College of Education and Human Development at Texas A&M University, College Station, Texas. Hersh C. Waxman is director of the Education Research Center and a professor in the Department of Teaching, Learning, and Culture of the College of Education and Human Development at Texas A&M University, College Station, Texas. Danielle B. Brown is an assistant research professor in the Department of Teaching and Learning of the College of Education at the University of Nevada, Las Vegas. Larry J. Kelly is a clinical professor in the Department of Teaching, Learning, and Culture and the Reta Haynes Clinical Faculty Fellow in the College of Education and Human Development at Texas A&M University, Coilege Station, Texas. ndweber@tamu.edu, hwaxman@tamu.edu, Danielle.b.brown@unlv.edu, & lkelly@tamu.edu

the development of statewide longitudinal data sets to assess teacher effectiveness. Recognizing that simply comparing teachers' standardized test scores, regardless of the student population, does not provide a valid assessment of a specific school's or teacher's effect on student learning, methods such as value-added modeling (VAM) have emerged in an effort to estimate teacher quality based on student improvement from year to year (Doran & Fleischman, 2005). This focus on individual teachers combats the common assumption that teacher effectiveness is consistent across classrooms within a particular school, while neglecting to appreciate the impact of each individual educator, also known as the widget effect (Weisberg, Sexton, Mulhern, & Keeling, 2009). Yet even the most sophisticated VAM techniques do not uncover what actually goes on in effective teachers' classrooms. Only direct classroom observations can reveal the subtle nuances and dynamic intricacies of effective teaching (Kane, Taylor, Tyler, & Wooten, 2011).

Observation research is a valuable method for studying classroom contexts because it allows researchers to collect detailed information about environmental characteristics and student and teacher behaviors within natural and authentic settings. It has been widely used to collect data with respect to student–teacher interactions (Pianta, la Paro, Payne, Cox, & Bradley, 2002), technology integration (Inan, Lowther, Ross, & Strahl, 2010), instructional quality (Stuhlman & Pianta, 2009), and specific teaching and learning behaviors (Waxman, Padrón, Franco-Fuenmayor, & Huang, 2009).

Classroom observation protocols are unique, as they focus on the aspects of teaching that can be reliably observed and assessed (Hamre et al., 2013) for the purpose of describing teachers' instructional practices (Ross, Smith, Alberg, & Lowther, 2004). The data collected from such measures directly inform the improvement of teaching practices (Hill & Grossman, 2013; New Teacher Project, 2013; Ross et al., 2004) based on what is determined to be effective (O'Leary, 2012; Taylor & Tyler, 2012). Furthermore, observations can be triangulated with other data, such as student achievement scores and survey responses, to identify specific teaching practices that lead to positive student outcomes (Raphael, Pressley, & Hohan, 2008), such as learner engagement (Raphael et al., 2008; Ross et al., 2004) and academic achievement (Kane et al., 2011). The incorporation of observation into the evaluation of teaching practices supports our overall understanding of effective teaching (Waxman et al., 2009) and directly responds to NCLB and Race to the Top's push for data-driven practice by allowing for the examination of how those teaching practices relate to student achievement.

Classroom Observations and Teacher Evaluation

Stemming from the national emphasis on academic standards and quality teaching, classroom observations are commonly used as an evidentiary basis for assessing teachers in the field (Kane et al., 2011; New Teacher Project, 2013;

O'Leary, 2012) and as a method for holding them accountable for student learning (Hamre et al., 2013). Meaningful feedback gathered from observational tools encourages both new and experienced teachers to improve their practice (Kane & Staiger, 2012) while offering administrators strong evidence to guide instructional and personnel-related decisions. Of particular interest is the potential for classroom observations to overcome the limitations of the value-added approach to teacher evaluation (e.g., some courses and grade levels are not tested, and some assessments are not designed to measure student growth) to evaluate teacher quality. Classroom observations measure teaching practices and enable the researcher to establish relationships between ratings and student learning (Sartain et al., 2011; Stuhlman & Pianta, 2009).

Classroom Observations and Teacher Education

Within the context of teacher education and preparation, observation practices are often implemented as a program requirement. Candidates are required to observe experienced educators, who serve as models of effective teaching practice. Previous studies have examined how these observational experiences develop an understanding of teaching and learning processes (Starks, Nicholas, & Macdonald, 2012) and of pedagogical content knowledge (Xiong, 2013), in addition to how their benefits are affected by method and type of observation (i.e., on-site vs. video; Pickering & Walsh, 2011).

In addition to content knowledge and candidate quality, clinical, field-based experiences are crucial for future teachers (Learning, 2010). Simply learning about teaching strategies and curriculum in course work is insufficient (Zeichner, 2010); candidates must be given opportunities to apply their knowledge in authentic settings, demonstrating that they can bridge the gap between theory and practice and develop a deeper understanding of the classroom environment (Darling-Hammond, 2006; Snyder, 2012). Teacher educators use a wide range of clinical practice models to develop candidates' pedagogical skills (e.g., student teaching; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Greenberg, Pomerance, & Walsh, 2011; coteaching; Von Zastrow, 2009; urban teacher residencies; Berry, Montgomery, & Snyder, 2008; Newman, 2009; Papay, West, Fullerton, & Kane 2012; internships; O'Brien, 2010), but, regardless of the model, teacher education field experiences must provide candidates the opportunity to measure their own success and effectiveness based on student learning outcomes (Snyder, 2012). The experiences of being observed in the classroom and receiving feedback from trained observers can directly facilitate this type of reflection and consequent growth, which is needed for preservice and early-career teachers to reach their potential.

Hundreds of research studies, policy analyses, and anecdotal reports have documented the challenges beginning teachers face (e.g., Veenman, 1984). As they try to keep up with planning and grading loads, manage their classrooms,

and navigate the responsibilities both in and out of the classroom that come with being an educator, early-career teachers transition from "survival mode" to tentative confidence when they begin to turn their attention toward developing their pedagogical skills and growing their toolboxes of teaching methods (Vonk, 1989). Preservice classroom teaching experiences clearly facilitate the growth of important pedagogical skills while easing the transition from candidate to teacher and the development of a professional identity. However, the increased interest in clinical experiences in teacher education has not led to the development of sensitive tools for evaluating these experiences. Existing observation tools have utilized rating scales or checklists rather than systematic observations (Kane, Kerr, & Pianta, 2014; Waxman, Weber, Franco-Fuenmayor, & Rollins, 2015). Furthermore, the tools currently available for classroom observation may not be appropriate for the study of teacher education programming and component outcomes. This potential shortcoming underscores the role that observational data should play in teacher education program development and evaluation. To address this need, the present study tested three existing observation instruments for the purpose of gathering classroom-level data for two distinct groups: (a) teaching candidates engaged in their final clinical field experience as full-responsibility teaching interns and (b) more experienced teachers. By comparing the observed behaviors, interactions, engagement with students, and classroom environments of the two groups, we were able to better understand how the teaching practices of novice teachers differ from those of more experienced educators. The knowledge gleaned from this study can be used to refine teacher education practices to better prepare novices for the realities of teaching.

The purpose of the present study is to examine how first-year secondary teaching interns' classrooms compare to those of more experienced teachers. Through the simultaneous use of three unique observation instruments, we addressed the following research questions: (a) How do first-year secondary teachers' classroom behaviors compare to those of more experienced teachers? (b) How do first-year secondary teachers' students' behaviors compare to those of more experienced teachers? and (c) How do first-year teachers' overall classroom environments compare to those of more experienced teachers? Each of the instruments revealed a different perspective of the classroom procedures and combined to provide a comprehensive picture that was not otherwise possible through use of any one instrument alone.

Methods

Participants

The internship program group consisted of 18 first-year secondary teachers in a field-based internship program that was part of their MEd course work at a large, research-based university in Texas. The internship positions were located at

a variety of middle and high school campuses in both rural and urban areas across Texas. The observations took place during the spring semester, and participants were notified within a week prior to the observations. The group consisted of a stratified random sample of teachers from the program who were teaching within a 100-mile radius of the university.

The comparison group consisted of teachers with approximately 8 years of successful classroom experience who had attended various teacher education programs. All participants in this group volunteered to participate in the study. The 18 members of the comparison group who were included in the study were matched to the intern group according to grade level and content area taught as well as by general school characteristics.

To ensure the validity of the matched samples and the comparison between the internship group and the comparison group, campus makeup information was obtained from the Texas Education Agency's Academic Excellence Indicator System campus reports. Based on the most recent available data, the 2011–2012 reports, an analysis of variance showed that there were no statistically significant differences between the schools where the two groups of participants taught in terms of percentages of economically disadvantaged, limited English proficiency, at-risk, African American, Hispanic, White, and Asian students.

All of the internship group cases were matched with cases from the comparison group. The participants in both groups of the study consisted of the teachers for each of the selected classrooms and three to five students from each classroom. The observed students were randomly chosen in each class by the observer at the beginning of the observation class period (~ 50 minutes) in an effort to closely represent the gender and age makeup of the group. Names and any other identifying information were not collected to preserve the anonymity of the students. The classes ranged from 8th to 12th grade, and the content areas included mathematics, science, social studies, language arts, and foreign language courses.

Instruments

Three different descriptive instruments were used during the observations to collect data about the teachers, the students, and the overall classroom environments.

Teacher observation instrument. The teacher observation instrument was adapted from the Teacher Roles Observation Schedule (Waxman, Wang, Lindvall, & Anderson, 1988) for the authors' purposes. It consisted of behaviors and characteristics in the following categories: interactions (e.g., with student(s)–instructional, with student(s)–managerial), setting (e.g., whole class, individual), instructional orientation (e.g., direct instruction, seatwork), nature of interaction (e.g., questioning, explaining), purpose of interaction (e.g., focus on content, redirect student thinking), and instructional technology (e.g., to present material, as a communication tool). At the end of each 30-second observation cycle, the observer checked off each

observed characteristic or activity. At the conclusion of the observed class period, percentages were calculated for each based on how many times it was observed out of the total number of cycles. The mean interrater agreement across all observers was high (.94).

Student observation instrument. The student observation instrument was adapted from the Student Behavior Observation Schedule (COS; Waxman et al., 1988) for the authors' purposes. It included characteristics and activities in the following areas: classroom setting (e.g., whole class, individual), manner (on- or off-task), types of engagement (behavioral, cognitive, and affective), interaction (e.g., with teacher–instructional, with other students), activity types (e.g., written assignment, questioning, distracted), educational use of technology (e.g., gather information, word processing), and technology (interactive whiteboard, desktop computer). At the end of each 30-second observation cycle, the observer checked off each observed characteristic or activity. At the conclusion of the observed class period, percentages were calculated for each based on how many times it was observed out of the total number of cycles. The mean interrater agreement across all observers was high (.97).

Overall classroom observation instrument. The overall classroom observation instrument was adapted from Part 4 of the Classroom Observation Measure (Ross & Smith, 1996) for the authors' purposes. The instrument addressed behaviors of the teachers and students as well as characteristics of the classroom environment. At the closing of each observation, the observer utilized the instrument by marking the degree to which each behavior and characteristic was observed (1 = "not observed at all," 2 = "some extent [once or twice]," or 3 = "great extent [3 or more times]"). The mean interrater agreement across all observers was high (.89).

Data Collection and Analysis

All observers who collected data for either group were trained to use each of the three instruments in classroom settings, and Cohen's kappa and interrater reliability showed that all results are reliable. For both groups, observation data were systematically collected by one of seven trained observers over the course of single secondary class periods. The teacher and between three and five students in each classroom were observed by way of time sampling in cycles for 30-second intervals. The number of cycles ranged from 5 to 10, depending on the length of the classes. For each cycle, the observed characteristics and behaviors were checked off, and at the end of the class periods, the observer calculated and documented the percentage of the sampled time that each of those characteristics and behaviors were observed for the individual participants. The observer immediately completed the overall classroom and College and Career Readiness Standards instruments at the end of each observed class period. All classrooms observed in both groups

were focused on content-related lessons that were designed to address specific state-guided curriculum standards.

Multivariate analysis of variance (MANOVA) and follow-up univariate tests were used to analyze the observation data. Each section of each of the three instruments (e.g., nature of interaction, student activity types, and teacher instructional behavior) was independently analyzed. The teacher and student observation analyses were based on the percentage of class time occupied by the specified behaviors, actions, and interactions. The overall classroom analyses addressed the extent to which the teacher and student instructional practices and the classroom environment characteristics were observed (i.e., not observed at all, to some extent, or to a great extent).

Results

Teacher Observation

Table 1 reports the overall findings from the teacher observations. In the internship program classrooms, the predominant setting or context observed was whole-class instruction (59.45%), followed by individualized work (26.67%) and, finally, small-group instruction (7.78%) and dyads (6.47%). In these settings, direct instruction took place about 46.67% of the time, instruction was learner centered 34.44% of the time, and students participated in seatwork 17.78% of the time. The teachers interacted with their students in an instructional context (58.89%), in a managerial context (27.22%), collaboratively (10%), and in a social way (5.56%). The nature of these interactions most often involved explanation (58.33%), cueing or prompting (49.44%), and questioning (32.22%), with the purpose of focusing on content (62.78 %) or work product (20%) and connecting content to real-life issues (18.33%). Instructional technology was used approximately 50% of the time, and most often with the purpose of presenting material (38.33%). It should be pointed out that the standard deviations are quite large across observed teacher behaviors and characteristics, with greater variability among first-year teachers than among more experienced teachers.

In the comparison group classrooms, the predominant setting or context observed was whole-class instruction (48.68%), followed by small-group instruction (29.62%) and, finally, individualized work (15.03%) and dyads (6.11%). In these settings, learner-centered instruction took place about 49.63% of the time, direct instruction occurred 41.91% of the time, and students participated in seatwork 5.06% of the time. The teachers interacted with their students in an instructional context (77.87%) and in a managerial context (14.33%). They did not interact with their students at all 6.56% of the time. The nature of the interactions most often involved explanation (69.01%), questioning (40.98%), and cueing or prompting (20.83%), with the purpose of focusing on content (67.88%) or work product (29.89%) and

Table 1

MANOVA and Analysis of Variance (ANOVA) Results for Teacher Behaviors and Interactions

	MANOVA		Intern group[a]		Comparison group[a]		
	df	F	M	SD	M	SD	ANOVA, F
Interactions	5	2.60*					
No interaction			2.78	8.26	6.56	11.35	
Instructional			58.89	26.10	77.87	20.86	5.78*
Managerial			27.22	16.74	14.33	14.27	6.18*
Social/personal			5.56	15.04	1.79	5.61	
Collaborative			10.00	20.86	1.11	4.71	
Setting	5	1.97					
Whole class			59.45	34.38	48.68	26.22	
Small group (>2 students)			7.78	20.74	29.62	27.09	
Dyads (2 students)			6.47	16.18	6.11	18.52	
Individual			26.67	32.90	15.03	20.81	
Traveling			0.00	0.00	0.56	2.36	
Instructional orientation	4	1.27					
Direct instruction			44.67	35.65	41.91	28.20	
Seatwork			17.78	22.64	5.06	16.56	
Learner centered			34.44	34.17	49.63	29.66	
Other			3.33	7.67	3.39	8.46	
Nature of interaction	9	3.76**					
Questioning			32.22	29.01	40.98	28.49	
Explaining			58.33	27.06	69.01	22.06	
Positive commenting			4.44	8.55	9.23	7.73	
Negative commenting			0.00	0.00	1.17	3.42	
Neutral commenting			5.56	15.04	3.50	6.46	
Listening			3.33	7.67	16.80	16.44	9.92**
Cueing or prompting			49.44	35.39	20.83	28.71	
Modeling/demonstrating			15.00	26.18	12.41	16.87	
Other			5.56	11.49	9.11	13.55	
Purpose of interaction	19	1.86					
Focus on content			62.78	27.40	67.88	27.94	
Focus on process			18.33	22.29	17.84	28.38	
Focus on work product			20.00	14.14	29.89	23.40	
Connect content to other disciplines			1.11	4.71	0.00	0.00	
Connect content to real-life issues			18.33	22.29	24.32	33.01	
Redirect student thinking			2.22	6.47	17.67	19.13	
Show interest in student work			8.89	17.11	9.64	11.61	
Show personal regard for student			5.56	15.03	2.66	5.29	
Encourage students to help each other			2.22	6.47	2.78	9.58	
Encourage students to succeed			11.76	15.90	5.18	9.16	
Encourage students to question			0.00	0.00	3.89	9.79	
Encourage extended responses			8.89	15.68	16.05	22.31	
Encourage self-management			17.22	11.79	5.68	8.79	
Praise student behavior			1.11	4.71	0.00	0.00	
Correct student behavior			13.33	16.80	2.96	8.23	
Correct student performance			0.00	0.00	5.99	10.73	
Assess prior knowledge			11.11	15.68	14.44	28.12	
Assess new knowledge			1.11	4.71	0.44	1.89	
Other			2.22	6.47	6.44	12.99	
Instructional technology	5	4.36**					
Use tech to present material			38.33	33.30	15.57	19.85	6.21*
Assist students with tech			7.78	20.74	1.11	3.23	
Use tech as a communication tool			2.78	11.79	16.40	28.78	
Use tech to create			0.00	0.00	0.56	2.36	
Use tech to access the Internet			1.67	5.14	4.44	9.22	

[a]n = 18; ***$p < .001$. **$p < .01$. *$p < .05$.

connecting content to real-life issues (24.32%). Instructional technology was used approximately 38% of the time, most often as a communication tool (38.33%) or to present material (15.57%). It should be pointed out that the standard deviations are quite large across observed teacher behaviors and characteristics, with greater variability among individual teachers in the comparison group.

The MANOVA results reveal a significant multivariate effect for the project (i.e., internship group vs. comparison group) on the Interaction, Nature of Interaction, and Instructional Technology sections of the teacher observation instrument. Follow-up univariate tests revealed that the internship group was observed significantly more often to be (a) interacting with students in a managerial way and (b) using technology to present material than teachers in the comparison group. Conversely, teachers from the comparison group were observed (a) interacting with students in an instructional way and (b) listening significantly more often than the intern group.

Student Observation

Table 2 reports the overall findings from the student observations. In internship group classrooms, the predominant setting or context observed was whole-class instruction (53.1%), followed by individualized or independent work (26.21%) and small-group instruction (12.87%). In these settings, students interacted with their teachers in either an instructional or a managerial context 11.27% of the time and with others (e.g., students) 21.61% of the time. The most prevalent activity that students were observed doing was watching or listening (41.49%). The next most prevalent activities were working on written assignments (35.06%) and reading (27.01%). Students were observed being on-task 77.01% of the time when they were engaged behaviorally (45.75%) or cognitively (34.26%). Interactive white-boards were used 10.92% of the time, often for gathering information (17.01%). The standard deviations vary widely across the observed student behaviors for the internship group.

In the comparison group classrooms, the predominant setting or context observed was whole-class instruction (49.48%), followed small-group work (26.55%) and individual instruction (13.97%). In these settings, students interacted with their teachers in either an instructional or a managerial context 20.27% of the time and with others (e.g., students) 26.14% of the time. The most prevalent activity that students were observed doing was listening or watching (47.64%). The next most prevalent activities were working on written assignments (65.10%) and discussing (25.43%). Students were observed being on-task 86.90% of the time when they were engaged behaviorally (59.05%) or cognitively (26.91%). Laptop computers were used 18.60% of the time, often for gathering information (8.20%). The standard deviations vary widely across the observed student behaviors for the comparison group.

The MANOVA results reveal a significant multivariate effect for the project (i.e., internship group vs. comparison group) on all sections of the student observation

Table 2

MANOVA and ANOVA Results for Student Behaviors and Interactions

	MANOVA		Intern group[a]		Comparison group[b]		
	df	F	M	SD	M	SD	ANOVA, F
Setting	5	2.70*					
Whole class			53.10	36.10	49.48	32.33	
Small group (>2 students)			12.87	28.77	26.55	31.05	7.59**
Dyads (2 students)			6.44	17.85	8.57	18.55	
Individual			26.21	32.47	13.97	24.91	6.13*
Other			0.00	0.00	1.85	12.88	
Manner	2	3.24*					
On-task			77.01	28.25	86.90	20.48	5.46*
Off-task			17.95	24.58	12.93	20.58	
Types of engagement	3	2.89*					
Behavioral (active response)			45.75	25.68	59.05	36.07	6.87**
Cognitive (expending mental effort)			34.26	24.09	26.91	35.92	
Affective (emotional reaction)			0.23	2.14	0.88	3.49	
Interactions	5	2.76*					
No interaction			67.13	28.40	52.27	33.47	8.46*
With teacher (instructional)			9.20	15.42	13.04	19.22	
With teacher (managerial)			2.07	6.13	7.23	18.02	6.14*
With other students			21.61	24.39	26.14	27.03	
Other			.023	2.14	0.47	2.48	
Activity types	16	2.27**					
Written assignment			35.06	25.28	35.10	28.75	
Assessments			2.30	9.49	1.18	10.30	
Discussing			11.95	26.80	25.43	28.75	8.54**
Reading			27.01	29.69	15.67	21.01	6.58*
Tutoring			0.00	0.00	0.00	0.00	
Working kinesthetically			0.92	4.21	0.51	2.26	
Answering teacher-posed questions			3.56	10.23	11.81	22.09	9.31*
Answering peer-posed questions			1.38	5.32	2.13	7.55	
Questioning			3.10	7.20	4.78	8.11	
Presenting			0.00	0.00	0.18	1.42	
Exploration/inquiry			5.06	14.38	15.81	25.81	10.49***
Using concrete learning materials			8.15	18.40	12.42	21.75	
Listening/watching			41.19	30.56	47.64	29.13	
Distracted			20.00	25.38	13.09	19.88	
Acting out (behavior)			0.69	4.77	0.71	3.29	
No activity/transition			2.99	8.09	2.13	4.53	
Other			4.48	11.98	4.32	11.33	
Educational use of tech.	6	4.99***					
Basic skills/drill/practice			1.38	5.32	2.72	11.38	
Gather information			17.01	23.33	8.20	14.96	6.74**
Organizing/managing/analyzing info			0.69	4.77	4.23	9.22	9.27**
Communicating/displaying findings			0.00	0.00	2.95	9.33	8.74**
Word processing			0.00	0.00	1.64	12.80	
Other			10.99	21.03	7.47	22.76	
Technology	5	16.75***					
Interactive whiteboard			10.92	22.55	4.10	18.20	
Laptop computer			0.00	0.00	18.60	30.19	33.14***
Desktop computer			2.30	15.07	0.00	0.00	
Other			19.89	21.21	19.79	30.80	
Other			0.46	3.01	14.04	23.71	27.97***

[a]n = 87. [b]n = 61.
***$p < .001$. **$p < .01$. *$p < .05$.

instrument, including setting, manner, types of engagement, interactions, activity type, educational use of technology, and technology. Follow-up univariate tests revealed that there were significant differences between internship and comparison group classes on the variables of small-group and individual settings; on-task manner; behavioral engagement; no interaction; managerial interaction with the teacher; discussing; reading; answering teacher-posed questions; exploration or inquiry; using technology to gather information, organize/manage/analyze information, and communicate and display findings; and laptop use. Students from the internship group classes were observed significantly more often (a) working in an individualized setting, (b) not interacting, (c) reading, and (d) gathering information with technology. Conversely, students from comparison group classes were observed significantly more than students from the effective school (a) in a small-group setting, (b) on-task, (c) behaviorally engaged, (d) interacting with the teacher in a managerial context, (e) discussing, (f) answering teacher-posed questions, (g) exploring or inquiring, (h) organizing, managing, and analyzing information, (i) communicating and displaying findings, and (j) using laptop computers.

Overall Classroom Observation

Table 3 reports the overall findings from the classroom observations. In internship group classrooms, the instructional behaviors of the teachers that were observed to the greatest extent included providing feedback (2.72/3), having warm and supportive relationships with students (2.56/3), acting as a coach or facilitator (2.50/3), providing opportunities for problem solving (2.50/3), and asking open-ended questions (2.50/3). The most widely observed student behaviors included engaging in classroom activities (3.00/3), asking questions indicating reflection (2.44/3), taking responsibility or ownership of work (2.39/3), and participating in learner-centered activities (2.39/3). The most commonly noted characteristic of the classroom environment was that the transitions were quick and efficient (2.17/3). The standard deviations for all but two of the variables were less than 1, suggesting there is a relatively small variance among overall environmental characteristics from the internship group classrooms.

In comparison group classrooms, the instructional behaviors of the teachers that were observed to the greatest extent included having warm and supportive relationships with students (2.89/3), sharing intellectual control with students (2.83/3), providing feedback (2.83/3), creating occasions for students to work out content (2.78/3), and distributing feedback evenly (2.67/3). The most widely observed student behaviors included taking responsibility and ownership of work (2.83/3), engaging in classroom activity (2.78/3), participating in learner-centered activities (2.67/3), and offering and defending prior views (2.06/3). The most commonly noted characteristics of the classroom environment were that the transitions were quick and efficient (2.33/3) and that materials and/or manipulatives were available for

Table 3

MANOVA and ANOVA Results for Overall Classroom Environment

	MANOVA		Intern group[a]		Comparison group[a]		
	df	F	M	SD	M	SD	ANOVA, F
Instruction (teacher)	29	3.49					
Shared intellectual control with students			2.28	0.83	2.83	0.38	
Created occasions for students to work out content			2.06	1.00	2.78	0.55	
Provided choice and independent decision making			2.06	0.94	2.56	0.70	
Provided diverse ways to experience success			1.56	0.62	2.11	0.76	
Promoted talk that was exploratory, tentative, and hypothetical			2.22	0.88	2.28	0.75	
Encouraged students to learn from other students			1.38	0.79	1.83	0.92	
Built an environment that supported risk taking			2.28	0.83	2.11	0.76	
Used intellectually challenging teaching procedures			1.78	0.65	1.67	0.77	
Used teaching procedures designed to promote quality learning			2.33	0.69	2.11	0.83	
Developed students' awareness of the big picture			2.28	0.89	2.06	0.80	
Raised students' awareness of different aspects of quality learning			1.44	0.62	1.39	0.61	
Promoted assessment as part of the learning process			1.89	0.68	1.50	0.86	
Facilitated students' activities and encourage participation			2.33	0.69	2.50	0.62	
Linked concepts and activities together			2.44	0.62	1.94	0.64	
Applied new concepts to similar situations			1.94	0.80	2.00	0.77	
Acted as coach/facilitator			2.50	0.71	2.61	0.70	
Provided opportunities for problem solving			2.50	0.71	2.17	0.92	
Asked open-ended questions			2.50	0.71	2.56	0.70	
Provided feedback			2.72	0.46	2.83	0.51	
Provided wait time for student responses			2.11	0.83	2.33	0.77	
Integrated technology into the lesson			2.33	0.69	2.00	0.91	
Distributed feedback evenly			2.39	0.70	2.67	0.59	
Scaffolded/redirected student thinking			2.22	0.65	2.61	0.61	
Related concepts to real-world problems/solutions			2.33	0.77	2.33	0.77	
Used a variety of modalities			1.89	0.76	1.72	0.89	
Varied instructional styles			1.94	0.80	1.61	0.78	
Offered encouragement of students' efforts			2.33	0.59	2.22	0.81	
Had warm, supportive relationships with students			2.56	0.62	2.89	0.32	
Linked students' prior knowledge to the current lesson			2.39	0.70	2.61	0.61	
Student	21	2.52*					
Offered and defended prior views			1.72	0.83	2.06	0.87	
Took responsibility/ownership of work			2.39	0.78	2.83	0.50	
Challenged/questioned content			2.22	0.65	1.56	0.70	8.74*
Asked questions indicating reflection			2.44	0.70	2.00	0.69	
Connected ideas and concepts			2.28	0.67	2.00	0.69	
Used different ways to answer			1.50	0.71	1.50	0.71	
Used technology for problem solving/creativity			1.50	0.71	1.67	0.97	
Used technology to learn basic skills			1.28	0.67	1.17	0.51	
Used technology to access the Internet			1.28	0.67	1.33	0.77	
Engaged in classroom activity			3.00	2.54	2.78	0.43	
Activities were learner centered			2.39	0.70	2.67	0.69	
Solved problems using real-life objects in the classroom			1.50	0.86	1.17	0.51	
Engaged in activities that integrated multiple subject areas			1.39	0.50	1.50	0.86	
Freedom of movement and placement during activities			1.61	0.85	2.00	0.84	
Classroom arrangement/ environment	3	1.50					
Materials and/or manipulatives available for hands-on practice			1.72	0.96	2.33	0.91	
Student work was displayed			1.72	0.83	2.17	0.92	
Transitions were quick and efficient			2.17	0.62	2.33	0.69	
Technology was accessible for student use			1.72	0.89	2.06	1.02	

Note. $n = 18$. 1 = not observed at all; 2 = some extent (once or twice); 3 = great extent (3 or more times).
***$p < .001$. **$p < .01$. *$p < .05$.

practice (2.33/3). The standard deviations for all but one variable were less than 1, suggesting that there is a relatively small variance among the overall environmental characteristics from the comparison group classrooms.

The MANOVA results reveal a significant multivariate effect for the project (i.e., internship group vs. comparison group) on the Student section of the overall classroom observation instrument. Follow-up univariate tests revealed that there is a significant difference between internship and comparison group classrooms on the variable of challenged/questioned content, which was observed more often in the internship group classrooms.

Overall and across all three instruments, we found substantial variability within both groups for many of the observed behaviors and environmental aspects. This large variability may be due to school-related factors and student demographic characteristics that impact classroom instruction. Content-related differences may also account for the wide variation with groups.

Discussion

Berliner (2004) pointed out that novice teachers tend to be very literal and rigid in their interpretation of what it takes to be a successful teacher. As experience accumulates, they begin to develop pedagogical intuition and become more flexible and responsive to the needs of their own classrooms (Berliner, 2004). This is consistent with the findings of this study. Overall, the first-year teachers in the internship group were focused on maintaining control of the classroom by requiring individual activity as well as using projectors to present material in a more traditional role as the "sage on the stage." The limited student interactions organized by the novice teachers resulted in mostly independent learning activities, suggesting that novice teachers do not feel as prepared to manage work groups of various sizes (Melnick & Meister, 2008). Conversely, teachers in the comparison group were observed facilitating more student-centered classes with a diverse range of instructional and learning practices. Their students were more often found to be on-task and behaviorally engaged with their peers in discussions and small-group activities. These findings and existing research assert that experienced teachers feel more confident in their own abilities to deal with a variety of behavioral issues that might arise in the classroom (Melnick & Meister, 2008).

The more experienced teachers in the comparison group were more at ease with classroom management and utilized a larger and more diverse range of teaching and learning strategies, giving the students a greater amount of autonomy and control over their own learning. The implementation of student-centered approaches, such as working together in small groups on inquiry-based activities, suggests that with experience comes a greater understanding of what management tactics work and an expanded collection of instructional strategies. To address this disparity in teaching styles between novice and experienced teachers, and to ensure that first-year teach-

ers feel better prepared and more at ease, teacher training may need to focus more on classroom management and instructional technology integration techniques.

The simultaneous use of multiple observation tools to examine several facets of the classroom environment supplied a rich, multidimensional conceptualization of the student–teacher dynamics for both groups. In this case, the different systematic instruments substantiated and expounded on each other, validating their respective findings. By dividing the focus of the observations between the three areas, teachers, students, and classroom environments, we were able to gain deeper insight into the dynamics of the observed classrooms than would have been possible with a single instrument. By using multiple observation protocols to study the classrooms of both novice and experienced teachers, we were able to discern differences in pedagogy and classroom environment that would not have been evident via other data collection methods. It would be prudent to extend this line of research in the future by observing more teachers who fall into each category to build a more robust database with greater reliability.

To effectively prepare teaching candidates for a smooth transition into the profession, teacher education programs should provide the knowledge and nurture the skills and dispositions of successful experienced teachers (Melnick & Meister, 2008). To do so, we must gain an understanding of the gaps in these areas between novice and experienced teachers. By conducting multifaceted observations of both groups, we have taken steps toward developing that understanding.

References

Berliner, D. C. (2004). Describing the behavior and documenting the accomplishments of expert teachers. *Bulletin of Science, Technology, and Society, 24*(3), 200–212. doi:10.1177/0270467604265535

Berry, B., Montgomery, D., & Snyder, J. (2008). *Urban teacher residency models and institutes of higher education: Implications for teacher preparation*. Carrboro, NC: Center for Teaching Quality.

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440. doi:10.3102/0162373709353129

Darling-Hammond, L. (2006). *Powerful teacher education: Lessons from exemplary programs*. San Francisco, CA: Jossey-Bass.

Doran, H. C., & Fleischman, S. (2005). Challenges to value-added assessment. *Educational Leadership, 63*(3), 85–87.

Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States*. Washington, DC: National Council on Teacher Quality.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal, 113*(4), 461–487.

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and

opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371–401.

Inan, F. A., Lowther, D. L., Ross, S. M., & Strahl, D. (2010). Pattern of classroom activities during students. *Teaching and Teacher Education, 26*, 540–546. doi:10.1016/j.tate.2009.06.017

Kane, T. J., Kerr, K. A., & Pianta, R. C. (Eds.). (2014). *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. San Francisco, CA: Bill and Melinda Gates Foundation.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Evaluating teacher effectiveness: Can classroom observations identify practices that raise achievement? *Education Next, 11*(3), 54–60.

Learning, T. B. (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers*. Washington, DC: National Council for the Accreditation of Teacher Education.

Melnick, S. A., & Meister, D. G. (2008). A comparison of beginning and experienced teachers' concerns. *Educational Research Quarterly, 31*(3), 39–56.

Newman, K. (2009). Teacher training, tailor-made. *Education Next, 9*(2), 42–49.

New Teacher Project. (2013). *Fixing classroom observations: How Common Core will change the way we look at teaching*. New York, NY: Author.

O'Brien, A. (2010). *A clinical approach to teacher preparation: A conversation with teacher educator Elizabeth Bondy*. Retrieved from http://www.learningfirst.org/visionaries/ElizabethBondy

O'Leary, M. (2012). Exploring the role of lesson observation in the English education system: A review of methods, models, and meanings. *Professional Development in Education, 38*(5), 791–810. doi:10.1080/19415257.2012.693119

Papay, J. P., West, M. R., Fullerton, J. B., & Kane, T. J. (2012). Does an urban teacher residency increase student achievement? Early evidence from Boston. *Educational Evaluation and Policy Analysis, 34*(4), 413–434. doi:10.3102/0162373712454328

Pianta, R. C., la Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *Elementary School Journal, 102*(3), 225–238.

Pickering, L. E., & Walsh, E. J. (2011). Using videoconferencing technology to enhance classroom observation methodology for the instruction of preservice early childhood professionals. *Journal of Digital Learning in Teacher Education, 27*(3), 99–108.

Raphael, L. M., Pressley, M., & Hohan, L. (2008). Engaging instruction in middle school classrooms: An observational study of nine teachers. *Elementary School Journal, 109*(1), 61–81.

Ross, S. M., & Smith, L. J. (1996). *Classroom observation measure observer's manual*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Ross, S. M., Smith, L. J., Alberg, M., & Lowther, D. (2004). Using classroom observation as a research and formative evaluation tool in educational reform: The school observation measure. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 144–173). Cambridge, England: Cambridge University Press.

Sartain, L., Stoelinga, S. R., Brown, E. R., Luppescu, S., Matsko, K. K., Miller, F. K., & Glazer, D. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago, IL: University of Chicago Urban Education Institute, Consortium on Chicago School Research.

Snyder, C. (2012). Finding the "royal road" to learning to teach: Listening to novice teacher voices in order to improve the effectiveness of teacher education. *Teacher Education Quarterly, 39*(4), 33–53.

Starks, D., Nicholas, H., & Macdonald, D. (2012). Structured reflective communication as a meta-genre in teacher education: Creative uses of "critique" in a teacher education program. *Australian Journal of Teacher Education, 37*(3), 90–110.

Stuhlman, M. W., & Pianta, R. C. (2009). Profiles of educational quality in first grade. *Elementary School Journal, 109*(4), 323–342.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? Evidence of systematic growth in the effectiveness of midcareer teachers. *Education Next, 12*(4). Retrieved from http://educationnext.org/can-teacher-evaluation-improve-teaching/

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*(2), 143–178. doi:10.3102/00346543054002143

Vonk, J. H. C. (1989). Beginning teachers' professional development and its implications for teacher education and training. *Irish Journal of Education, 23*(1), 5–21.

Von Zastrow, C. (2009). *Breaking new ground in teacher education: A conversation with teacher educator Nancy Bacharach*. Retrieved from http://www.learningfirst.org/visionaries/NancyBacharach

Waxman, H. C., Padrón, Y. N., Franco-Fuenmayor, S. E., & Huang, S. L. (2009). Observing classroom instruction for Ells from student, teacher, and classroom perspectives. *TABE Journal, 11*(1), 63–95.

Waxman, H. C., Wang, M. C., Lindvall, C. M., & Anderson, K. A. (1988). *Classroom observation schedule technical manual* (rev. ed.). Philadelphia, PA: Temple University, Center for Research in Human Development and Education.

Waxman, H. C., Weber, N. D., Franco-Fuenmayor, S. E., & Rollins, K. B. (2015). Research-based approaches for identifying and assessing effective teaching practices: Challenges, new directions, and policy implications. In Y. Li & J. Hammer (Eds.), *Teaching at work* (pp. 9–27). Rotterdam, Netherlands: Sense.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York, NY: New Teacher Project.

Xiong, W. (2013). A potential approach to pre-service teachers' professional learning: The video analysis of the authentic classroom. *U.S.–China Education Review, 33*(3), 149–161.

Zeichner, K. (2010). Rethinking the connections between campus course and field experiences in college- and university-based teacher education. *Journal of Teacher Education, 61*(1–2), 89–99. doi:10.1177/0022487109347671