Using Standardized Observation to Conceptualize, Measure, and Improve Classroom Processes

Robert C. Pianta and Bridget K. Hamre

Many children spend more of their waking hours in classrooms than they do at home; within these settings they are exposed to experiences that for better or worse, intended and unintended, shape their development—they may learn to read, write, and think critically; they make friends and have to face the inevitable challenges of peer relationships; and they are oriented increasingly to become productive, independent members of a larger society. Interest is keen in the extent of these classroom effects, the methods of producing and reproducing them at various levels of scale, and understanding the mechanisms responsible for them. In fact, it could be argued that these interests in classroom effects are among the core focus of educational policy, practice, and research at the present time. The scale is staggering, with millions of classroom teachers working each day in public education settings, more than 200,000 of whom are new entrants to the profession each year, and with recent reports indicating 87 percent of whom leave the profession before a decade of experience (Pianta and Allen, 2008; Loeb and Béteille, 2008; Anderson, 2008).

The focus of this chapter is the consequences for policymaking of research on teachers’ classroom interactions with students, particularly standardized observations of those interactions. We address implications not only for education policy that aims to improve student achievement (such as No Child Left Behind), but also research policy. Our overall goal is to advance classroom interactions and processes as a legitimate focus of education policy (Tseng and Seidman, 2007; Shinn and Yoshikawa, 2008), based on what we know from research using standardized observations of those processes.

To preview, we argue that advances in theory, in measurement, and in intervention have led to the possibility that policy can reach into classrooms more directly than it has heretofore, taking shape, for example, in metrics for a “highly effective teacher” that rely neither on the proxies of degrees or experience that bear only indirectly (Gordon, Kane, and Staiger, 2008) and often not at all (Pianta and Allen, 2008) on student outcomes nor on the tautology that effective teachers are those who produce achievement gains (Rivkin, Hanushek, and Kain, 2005; Rockoff,
which runs the risk of too narrowly defining an important social asset. Rather, we argue that it is now feasible to experiment with policy that focuses on direct assessments of teachers’ performance in the classroom as an instructor, socializer, motivator, and mentor while demonstrating explicit evidence that these assessments account for student learning gains.

We also argue that research could advance the impact of classrooms on student performance if standardized observations of teachers’ classroom practices were more routinely and strategically a focus or embedded in large-scale value-added studies.

Finally, we contend that placing observational assessment of teachers’ classroom instruction and interactions more squarely into the realm of policy and accountability frameworks would in turn spur research and development efforts in teacher training likely to lead to more regular and efficient production of effective teaching by training programs and through ongoing professional development supports that actually work. In the end, our interest is in how hiring, promotion, or reward structure in schools; certification and licensure at the state level; pre-service preparation of teachers; and funding of education science, if centered on performance-based observations of classroom teaching behaviors, could function in a coordinated fashion to more systematically and directly produce better teaching in classrooms.

We position this chapter as a direct response to years of research on education inputs and outputs that has been of two forms, which we argue are both lacking in their capacity to inform policies leading to improvement of classrooms: one form identifies connections between distal proxies for inputs and student test score outputs through large-scale econometric modeling of the “education production function,” but is uninformative as to the proximal mechanisms of production; and the second form produces rich descriptions of classroom processes using in-depth studies of cases, but has little bearing at much higher levels of scale (see Ladd, 2008). The models and methodologies for conceptualizing, measuring, and improving classrooms that rely on standardized, scalable, valid observational assessments of teacher performance, we argue, provide tools that complement econometric and case methods and hold promise for improving the effectiveness of American education.

**CLASSROOMS ARE THE LOCUS OF EDUCATION EFFECTS**

The last decade has witnessed a marked increase in research examining the complex nature of children’s experiences in classrooms and the ways in which these experiences uniquely contribute to children’s social, cognitive, and academic development. This increase in research has been driven by several factors. First, educational reform in general, accountability frameworks at the state level, and the No Child Left Behind Act in particular, have placed individual schools and teachers in the spotlight: requiring evidence of their role in producing student achievement. Thus, the attention of policymakers and school administrators has turned to the (admittedly too-thin) research base on links between classroom experiences and student outcomes.

This need for information on what happens in classrooms is exacerbated further as the results of widespread accountability and state standards-testing become used for identifying schools that do not meet established proficiency levels and rates of
performance. Witness the concern in any local paper or in state government when schools are not making “Adequate Yearly Progress” or fail to make accreditation standards. And reauthorization of NCLB is launching a range of discussions that shift the dynamic of accountability to focus on how inputs produce achievement through debate about effective or qualified teachers (e.g., Appalachian Regional Advisory Committee, 2005; Foundation for Child Development, 2005; Gordon et al., 2008; Hamre, Pianta, Mashburn, and Downer, 2007; Ladd, 2008; Perie, Moran, and Lutkus, 2005). Classrooms are the hotspot as states and districts scramble to find the right mix of curriculum, professional development, and instructional supports that will raise students’ achievement (e.g., Appalachian Regional Advisory Committee, 2005; Educational Testing Service, 2004; Lasley, Siedentop, and Yinger, 2006).

A second reason for focusing on classrooms is illustrated in the widespread and persistent arguments being made for state-supported pre-kindergarten as a means of addressing a range of social and economic concerns (Bogard and Takanishi, 2005). This evidentiary stream clearly demonstrates (more so than most effects-focused research) that classroom processes are the locus of early education program effects for high-risk young children. A host of studies describing results of randomized controlled trials and highly controlled quasi-experimental work, have shown quite clearly that significant gains in achievement accrue as a function of enrollment in pre-kindergarten (pre-k) classrooms, and that children’s interactions with teachers are in large part the source of these effects (see Mashburn et al., 2008).

We focus here in pre-k not as an end in itself, but illustratively; to the extent there is a body of well-controlled studies of educational effects of classrooms processes, it exists for pre-kindergarten educational programs. More widely, when schooling is viewed as a solution to address student performance gaps, social or economic inequality, or international competitiveness (Carroll, Fulton, Abercrombie, and Yoon, 2004), the rhetoric, evidence, and programs are focused on what accrues to students as a function of their engagement in classroom processes.

Not surprisingly, the third body of evidence in support of classrooms as a focus of policy draws directly from studies that indicate quite clearly that classrooms, and teachers, matter. In studies of large-scale statewide testing programs in which multilevel analysis has been used to isolate sources of variance to which achievement growth can be attributed (e.g., Nye, Kostanpoulos, and Hedges, 2004), classrooms are more often than not the greatest source of variation in what students learn and gain as a function of attending school (at least in achievement-related domains; it is not as clear with regard to social outcomes). Recent work, largely motivated by NCLB’s focus on highly qualified teachers and an assortment of concerns related to teacher licensing systems, has focused on teacher characteristics such as degrees, experience, qualifications, and test performance, and shown significant returns to achievement as a function of each of these features—and policy-relevant effect sizes when considering their cumulative benefits (see Ladd, 2008; Loeb and Béteille, 2008).

Classroom processes are implicated as significant moderators of treatment effects in highly controlled experimental work (Battistich, Watson, Solomon, Lewis, and Schaps, 1999). Within classrooms, differences in teachers’ implementation of treatments (e.g., curricula) appear to be the single biggest factor determining
effects on child outcomes (Domitrovich and Greenberg, 2004), even when teachers are instructed to deliver the intervention in a standardized manner and are provided with regular and intensive supports to maintain fidelity. Thus, in both value-added research and experimental studies not only is a good part of the value of attending school conveyed at the level of classrooms, but it also appears that teachers play a major role in determining the value of the classroom environment for student learning and development.

Thus, as the focus of research seeking to locate and maximize the benefits of education shifts to recognize and consider processes within classrooms that could account for such benefits, whether in value-added or experimental frameworks, the evidence shows that experiences with teachers can matter. Relatedly, new investments of research dollars and a targeted effort to engage researchers from other disciplines in the educational sciences have spurred interest in pre–k-12 classrooms, particularly among developmental psychologists. Thus, we now see the bridging of developmental and education science (Pianta, 2007) in ways that have produced marked gains in understanding classroom processes related to instruction, teacher-child interaction, and peer relations.

One thread of this connection can be traced to economists, who have been instrumental in advancing the science of value-added modeling of achievement gains, using longitudinal data to isolate effects of classroom-level experience for which, it is argued, causal influence can be attributed (see Ladd, 2008; Murnane and Steel, 2007). A second thread leading to interest in classrooms is evident when basic science disciplines, such as developmental psychology, move from lab to setting; for example, scientists interested in peer relations have moved from experiments in the lab to careful study of peer relations in classroom and school settings (Gifford-Smith and Brownell, 2003). In the case of peer relations, such movement to classroom settings has led to an understanding of how teachers set norms for aggressive behavior that have consequences for peer victimization and learning (Henry, Guerra, Huesmann, Tolan, VanAcker, and Eron, 2000) and how teachers’ use of language and responses to social cues of students of varying gender or ethnicity have consequences for motivation, learning, and social relationships in the classroom (Weinstein, 2008).

For the remainder of this chapter, we focus on one source of classroom effects on achievement—teacher-student interactions—and the consequences when an evidence base is accrued regarding the measurement, production, and improvement of interactions that are effective in producing achievement gains.

One reason for this focus is the evidence regarding teachers as a primary locus for schooling effects. The other reason is that for policymaking, the production of teachers (and presumably teaching quality) is of real concern, is under considerable regulatory control at the state and federal levels, and has extraordinarily high stakes attached to its success or failure—as just one example, see Pianta et al. (in press), who reports that in the thousands of U.S. classrooms taught by certified teachers, only 25 percent provide a level of instructional or emotional support consistent with the production of learning gains. As another, refer to the report by Clotfelter, Ladd and Vigdor (2007) that certain profiles of teacher characteristics have negative effects on achievement for certain student groups.
We argue that policies that place direct assessment of actual teaching as a central feature of accountability frameworks and provisions for equity of educational opportunity are likely to accomplish several interlocking aims that in a coordinated fashion could result in substantial shifts in the nature and quality of instruction, socialization, mentoring, and tutelage that takes place in classrooms. First, these policies will result in better teaching in classrooms; second they will lead to improved student outcomes; and third they will drive a research and development enterprise that will soon result in a science of the production of teaching and teachers—something that simply does not now exist.

TEACHER-CHILD INTERACTIONS MATTER

A major question occupying the attention of education science and policy has been to determine the degree to which placement in a particular classroom accounts for variability in children’s academic performance. Numerous studies have indicated that: 1) a significant portion of variance in student learning is explained at the classroom level (see Nye et al., 2004); and 2) deflections in the trajectory of student learning across years can be attributed to their experiences in specific classrooms (see Ladd, 2008; Rivkin et al., 2007).

Although these studies have been important in laying a foundation for inquiry into classroom effects, they fail to articulate specific processes that may lead to student learning and positive social adjustment. The problems with this approach are highlighted in recent debates about teacher education and teacher education research. Hanushek’s (2002) definition of teacher quality—“Good teachers are ones who get large gains in student achievement for their classes; bad teachers are just the opposite” (p. 3)—as well as much of the research using the value-added paradigm (see Gordon et al., 2006, for overview) provide virtually no guidance to the development of a scientifically driven agenda aimed at evidenced-based ways to produce good teaching, either through improved in-service support or teacher education (Cochrane-Smith and Zeichner, 2005).

Of course, some have argued that value-added models could be used to identify effective and ineffective teachers, who could then be observed to detect the behaviors that differentiate them (Gordon et al., 2006). Although this is a promising pursuit, make no mistake that it is fraught with serious methodological problems, notably what Garmezy, three decades ago, termed the “retrospective fallacy” (Garmezy, 1977). In research on the prediction of psychopathology, Garmezy noted the trend to identify carefully matched individuals with and without a disorder and then compare those pairs on measures of potential interest. Garmezy was careful to note that this strategy is flawed because a) it will always find a difference on some indicator (even if there is no detectable conceptual reason for the difference), b) it is not possible to attribute cause to indicators on which there is a difference because of problems with selection bias, and c) these indicators always perform more poorly (if at all) as discriminators in predictive models involving the entire population. Current interest in value-added models as a way of discriminating successful and unsuccessful teachers for further study may suffer from the same fallacy in terms of being a path for identifying factors causal to teachers’ putatively effective behaviors (or student achievement).
We argue that, on its own, value-added work as a means of identifying and producing effective teachers is inherently limited, both conceptually and methodologically because it relies on retrospective identification of effectiveness and has no conceptual framework for hypothesis testing and experimentation about putative mechanisms linking inputs to outcomes (Pianta et al., in press). The more important challenge is to identify (perhaps with value-added models as one source) a set of conceptually sensible and empirically valid teaching practices, and assign students randomly to teachers exhibiting those practices; or to build programs (interventions) for training teachers to exhibit those behaviors and test their efficacy on teacher and student outcomes.

Defining and Understanding Classroom Environments

Consistent with developmental theories emphasizing the role of proximal process on development (Bronfenbrenner and Morris, 1998), we take the view that a key ingredient of any classroom environment is interaction between adults and students. This view of classroom environments excludes a focus on some aspects of classrooms that have been the focus of research, such as the availability of materials or curricula, or the number of students in the setting. However, it provides a broad, holistic view of the classroom environment that includes all types of teacher-student interactions, those that are socioemotional, organizational, and instructional in nature. As such, our view of classroom environments is inclusive of research focused on more discrete aspects of classrooms such as quality or effective teaching, learning environments, and student-teacher and peer relationships.

In school, children’s experiences in classrooms constitute the majority of the proximal processes responsible for achievement and other outcomes. Although students do spend time in other places (e.g., lunchrooms, playgrounds) the bulk of their day is contained within classroom walls and experiences at the classroom level appear to be most closely associated with student outcomes (Nye et al., 2004). School-level effects, such as school climate and culture, are distal factors, the effects of which are largely mediated through or that moderate classroom process effects. For example, when schools have a positive climate, in which faculty, staff, and students feel safe and supported, teachers and students are better able to recreate this environment in the classroom than if the school climate is marked by hostility and discord—thus, the effects of school climate may be mediated by interactions in the classroom (Deemer, 2004).

Similarly, findings from intervention studies provide strong evidence of substantial variability of implementation between classrooms (Jones, Brown, and Aber, 2008) even when the entire school is being directed to adhere to a standard intervention protocol. Mashburn (Mashburn and Pianta, 2008) argues that structural features of schooling or classrooms that are typically the levers engaged by policymakers (e.g., teacher education) do not exert direct effects on child outcomes but instead should be treated conceptually and analytically as moderators of the effects of interactions in the classroom setting.

To help organize the diverse literatures that might inform the task of describing teacher-student interactions in classrooms, Hamre and Pianta (2007) presented the CLASS Framework. The CLASS Framework is a theoretically driven and empirically supported conceptualization of classroom interactions organized into three major domains—Emotional Supports, Classroom Organization, and Instructional
Using Standardized Observation to Conceptualize, Measure, and Improve Classroom Processes • Robert C. Pianta and Bridget K. Hamre

Supports. Within each domain are a set of more specific dimensions of classroom interactions that are presumed to be important to students’ academic and/or social development (see Figure 1). The CLASS framework starts with an understanding of the nature and regulators of developmental change at a given period (Hamre and Pianta, 2007; Pianta and Allen, 2008), then applies that understanding in a developmentally informed analysis of classroom settings and teacher-child interactions, maps that understanding back onto the rich and deep literature on classroom teaching and educational effectiveness, and then organizes this analysis within a framework that could inform measurement. McCaslin and colleagues (2006) have taken a similar approach in measuring classroom-setting effects on student motivation. The result is a multilevel conceptual and measurement model that can be subjected to empirical tests and evaluation.

The CLASS framework is one of several descriptions of classroom environments or quality teaching put forth in the educational and developmental literatures (e.g., Brophy, 1999; Brophy and Good, 1986; Eccles and Roeser, 1999; Gage, 1978; Pressley, Roehrig et al., 2003; Soar and Soar, 1979). For example, Brophy (1999) describes 12 principles of effective teaching, including supportive classroom climates, opportunities to learn, curricular alignment, thoughtful discourse, scaffolding engagement, and achievement expectations, each of which are based on research findings and theories of teaching and learning. Others organize teachers’ practices into larger domains of teaching and classroom environments. Pressley and colleagues (2003) draw from their studies of effective teachers (e.g., Bogner, Raphael, and Pressley, 2002; Pressley, Allington, Wharton-McDonald, Block, and Morrow, 2001; Pressley, Roehig et al., 2001; Wharton-McDonald, Pressley, and Hampston, 1998) to suggest that effective teaching strategies can be organized into decisions regarding motivational atmosphere, classroom management, and curriculum and instruction. Similarly, Eccles and Roeser (1999) suggest that schooling is characterized by organizational, social, and instructional processes that help regulate children’s and adolescents’ development across cognitive, social-emotional, and behavioral domains.

The CLASS framework draws heavily from this theoretical and empirical work. There are, however, several distinctions between the CLASS Framework and other frameworks for studying classroom, one being that Hamre and Pianta (2007) propose a latent structure for organizing teaching behaviors that in most approaches are simply culled and categorized by type. The latent structure in the CLASS model poses explicit, testable hypotheses regarding the organization of meaningful patterns of behavior(s) that are tied to underlying developmental processes. This conceptualization is explained in Pianta and Allen (2008) as a way of thinking about classroom supports that begin with an understanding of the developmental salience of various inputs and behaviors. The CLASS model then starts with a developmental analysis of settings, which is different than thinking of classrooms as only serving educationally framed aims (e.g., achievement) and then seeking to define what is good teaching. Importantly for policy, one implication of the CLASS conceptualization of classrooms is that the desirable outcomes related to classroom experience are not limited to performance on achievement tests, but rather involve a range of competencies that could be valued or assessed, and which could mediate or moderate achievement gains.
In this structure, meaningful units of teacher-child interaction are organized by patterns, which in turn are the basis for identifiable and scalable dimensions of interaction. These dimensions are then organized into one of three broad domains of classroom supports. For example, the domain of emotional supports includes three dimensions: positive classroom climate, teacher sensitivity, and regard for student perspectives. Organizational support includes effective behavior management, productivity, and instructional learning formats. And the instructional support domain includes concept development, quality of feedback, and language modeling dimensions. The organization of these dimensions into these three broad domains has been tested and validated for grades pre-k to fifth and a somewhat different set of dimensions, within the same hypothesized three-domain structure, have been specified for grades 6-12. The dimensions included in the current pre-k through fifth-grade framework have received the most empirical support but are not exhaustive; there are likely be other dimensions that could fit within each domain, and as we suggest, dimensions may vary with developmental period or epoch.

Within each of these dimensions are posited a set of behavioral indicators reflective of that dimension, which in turn are operationalized at various anchor points on a 1-7 scale using specific behaviors and interaction patterns that can be reliably observed in a specified window of time. For example, positive classroom climate includes observable behavioral indicators such as the frequency and quality of teacher affective communications with students (smiles, positive verbal feedback) as well as the degree to which students appear to enjoy spending time with one another. This detailed, multilevel conceptualization of teacher-child interactions in classroom environments, moving back and forth between broad theoretically-based domains to very specific behavioral indicators, is a way of organizing and linking, under one framework, approaches to classroom assessment and observation that are driven by different methodological and theoretical considerations and heretofore have remained distinct. It also presents a set of testable hypotheses concerning the organization of behavior at varying levels of abstraction.

A key distinction between the CLASS framework and others is that there is empirical support for this organization of classroom processes that derives from large-scale studies of actual classrooms (Hamre et al., 2007). It is not trivial to emphasize that these features of the supporting evidence—that it is empirical in nature and drawn from many thousands of observations—enhance the extent to which this evidence can have some bearing on policy and make this evidence base somewhat unique. Drawing from a sample of over just under 4,000 preschool to fifth-grade classrooms that were a part of several large, national and regional studies, Hamre and colleagues (2007) first examined observational instruments and sorted observed dimensions into the domains described by the CLASS framework. They then used confirmatory factor analysis to examine the extent to which this organization of classroom interactions was consistent with actual observations in these settings and to test alternative organizational structures. Results suggested adequate fit of the three-factor model and that the fit of this model was superior to a one- or two-factor model. These findings provide evidence that the three-domain structure suggested by the CLASS framework fits the natural variation in classrooms.

A final distinction between the CLASS framework and other conceptualizations of classrooms is that it was developed to apply to teacher-student interactions in
Using Standardized Observation to Conceptualize, Measure, and Improve Classroom Processes • Robert C. Pianta and Bridget K. Hamre

classroom contexts across grades and across content areas, from preschool to high school, and thus differs from work that has focused on more discrete groupings of grades, or has a specific focus on teachers’ content knowledge or delivery of content, when the focus is itself on content (the contribution to achievement gains of content-neutral CLASS ratings, content-specific behaviors, or teacher content knowledge remain open empirical questions). There are three reasons why the CLASS framework has not included, in a direct manner, content in the descriptors or scales it applies to classroom interactions. First, the system was conceptualized within a developmental psychology framework that posits that the key engines of development are the properties embedded in interactions between children and the setting, and that properties of these regulatory systems (such as feedback or reciprocity) are what drive development (Sameroff, 1983; Roeser, Eccles, and Sameroff, 1998). Second, CLASS was designed to be applicable across as many classroom settings as possible and to have utility as an observational tool even when the content focus of the lesson or instruction changed in unpredictable ways. Third, as we have had to address questions related to content, we have worked with individuals to identify and specify what would be desired in a content-focused scale (say, one focused on math in fourth grade), a large proportion of the behavioral indicators are pedagogical or interactive in nature and not content-specific (e.g., presents lesson in an organized manner). We refer readers to Hamre and Pianta (2007) for a more detailed description of, and rationale for, the domains and dimensions represented in this system of assessment.

We remind readers that CLASS encodes teacher-child interactions in terms of global ratings on dimensions represented in terms of seven-point rating scales, with such judgments typically being made after 20 minutes of observation. There are an assortment of questions pertaining to CLASS and other standardized observational protocols in classrooms about decomposing variance attributable to factors such as raters, time of day, time of year, length of observational window, and unit of analysis, and others (Raudenbush and Sadoff, 2008). Each of these parameters is treated as error in models attempting to isolate effects of teachers, and so to the extent they can be modeled in analysis of teacher effects, the precision of prediction improves as does effect sizes of teacher-related variables (Raudenbush and Sadoff, 2008).

In a series of analyses that cannot be fully documented here, but are available on request, CLASS ratings from multiple studies across multiple grades were examined in relation to these basic generalizable, theory-informed questions. The evidence suggests that indeed rater-related variance, even after raters pass stringent tests for interrater agreement and reliability, are substantial in these global dimensions of classroom interaction, but that such effects are attenuated with increased numbers of raters within a given data collection system, and pale in comparison to time of day and window effects for frequency-related codes (Chomat-Mooney et al., 2008). Indeed the evidence suggests that CLASS does capture aspects of teacher-child interaction that are stable across a specific day, across days, across students, and across content area of instruction, thus providing a reasonable estimate of features of a teachers’ behavior that appear stably characteristic of her interactions with students.
Epidemiology and Change of Classroom Environments

It is stunning, given the importance of classroom settings as vehicles for the transmission of knowledge and skill in our system of education, that little-to-no population level data exist pertaining to exposure of children and adolescents to particular classroom practices that are either known to relate to academic success or failure, desired on the basis of certain policies or values, or even hypothetically expected to relate to outcomes. Although, as we describe below, there is evidence emerging for early education and elementary classrooms, in secondary classrooms there is no current work which provides national-level, observational data on these environments. Two large national studies conducted over the past ten years do provide some of the first “epidemiological” data on preschool to fifth grade U.S. classrooms (Early et al., 2005; NICHD ECCRN, 2002, 2005; Pianta et al., 2005; Pianta, La Paro, and Hamre, 2008).

Overall these studies suggest that the average child is exposed to moderate levels of emotional support and classroom organization, and fairly low levels of instructional support throughout preschool to elementary school (Early et al., 2005; NICHD ECCRN, 2002, 2005; Pianta et al., 2005, 2008). In general, teachers are fairly positive in their interactions with students and examples of teacher or student negativity are relatively rare (NICHD ECCRN 2002, 2005). However, these interactions between teachers and students appear to be fairly impersonal, with very few instances in which individual students have positive, one-to-one interactions with their teachers.

For example, in fifth-grade classrooms, positive, individual interactions with a teacher occurred in only 1 percent of observed intervals across a school day (Pianta et al., 2008). The typical student also has few interactions with teachers around behavior management issues in schools, either positive or negative (NICHD ECCRN, 2005). However, one clear indication of problems in classroom organization comes from consistent findings that students spend a great deal of their time in classrooms without being exposed to any learning activity at all, ranging from 42 percent of the time in preschool classrooms to 30 percent of time in fifth-grade classrooms. Evidence on the quality of instructional supports is particularly concerning, with consistent evidence that children across grades are unlikely to be exposed to high-quality supports such as concept development and feedback. For example, by a ratio of nearly 5 to 1, children in fifth-grade classrooms are exposed to instructional activities (across any content area) that are basic-skill focused, in contrast to a focus on analysis, inference, or synthesis of information (Pianta et al., 2008).

Most notable in these and other studies, however, is the high degree of variability in classroom quality. A typical school day for some students includes spending the majority of time engaged in productive instructional activities with caring and responsive adults who consistently provide feedback and challenge students to think critically. For others, a typical day consists of spending most of the time sitting around, watching the teacher deal with behavioral problems, and engaged in boring and rote instructional activities such as completing worksheets and spelling tests (Early et al., 2005; NICHD ECCRN, 2002, 2005; Pianta et al., 2005, 2008).
These problems of inconsistent exposure to high quality classrooms are compounded by clear evidence of inequity. Students coming from disadvantaged backgrounds are more likely than their peers to be exposed to poor quality (Hamre and Pianta, 2005; Pianta et al., 2005). Further troubling is evidence that even the student lucky enough to experience a high quality classroom one year is very unlikely to be systematically exposed to high quality over a period of years, even if they remain in the same school (NICHD ECCRN, 2005; Pianta et al., 2008), suggesting that school-level resources such as professional development supports and school climate are insufficient to ensure high quality classroom environments. Taken together these studies suggest that very few of the students who are in greatest need of high quality classroom experiences receive them and the few that do are unlikely to receive them consistently, making it unlikely that the positive effects will be sustained.

**Changes in Classroom Environments across Grades**

Another area in which very little is known concerns patterns of exposure to various classroom supports as children move classrooms from year to year. Most research that does exist in this area has looked specifically at transition periods, either from preschool to kindergarten, into middle school, or into high school (Anderman and Midley, 1997; Ferguson and Fraser, 1998; Eccles, Flanagan, Lord, and Midgley, 1996; Hamre et al., 2007; Rimm-Kaufman and Pianta, 2000). For example, work on the transition to kindergarten has focused on ways in which classrooms become more oriented toward learning and less oriented toward social development (Hamre et al., 2007).

Others have provided evidence of the shift in goal orientations and student-teacher interactions from elementary to middle school, with students typically rating teachers as more distant and less supportive of autonomy in middle school—just at the time when, developmentally, young adolescents may be most in need of those positive supports from teachers (Anderman and Midgley, 1997; Ferguson and Fraser, 1998; Eccles et al., 1996). A recent study provides evidence that this shift is reported by teachers as well; as children move from kindergarten through sixth grade, there is a general pattern of decreased relational connections (both positive and negative), particularly around fifth grade (Jerome, Hamre, and Pianta, in press). Finally, during the transition from middle to high school, students report decreased levels of engagement with the content of schooling (Yair, 2000). However, we know little about normative shifts outside of these crucial transition periods.

In sum, despite the importance of teacher-student interactions in classrooms to student achievement and other outcomes, and that such processes can be reliably assessed and have been shown to predict to achievement, only one major national study provides information on the nature and quality of these key educational assets for large numbers of students. Moreover, this study, although large, is not representative of national or state demographics and under-represents students of color or from families of varying language or economic backgrounds. In comparison, there are literally dozens of large-scale epidemiological studies of health-related inputs.
Improving Teacher-Student Interactions

A final set of issues to be addressed concerns how and the extent to which student-teacher interactions can be improved. By and large these issues concern the nature and effectiveness of professional development and certification. At present, there is a consensus that the vast machinery of teacher licensing, certification, entry-level and advanced degrees, and in-service professional development play far too little a role in producing achievement gains for students (or effective teacher-student interactions as evinced in the Pianta et al. study described above). The one exception to this pattern is the set of findings showing that teachers with degrees and courses in math tend to produce higher math scores for their students (see Goldhaber, 2008). Rather, the general conclusion to be drawn from this work is that the metrics and processes used in the training and certification of teachers are very hard to link, either directly or indirectly, to student achievement gains or to observed effectiveness of teachers’ interactions in classrooms. One argument that has been advanced as an explanation for this is that the metrics used to regulate, certify, license, and produce teacher quality are not tied to direct observation (measurement) of what teachers do.

As accountability frameworks penetrate teacher education programs in a variety of ways—data systems that link teacher preparation experiences to value-added modeling of student achievement, studies of program ‘output’ and there is a noticeable push for teacher education itself to be more strongly grounded in evidence and research. By nearly every visible marker, the pressure is on teacher education to deliver supports and training that can demonstrably produce more effective classroom learning environments for students. Critically, this movement of attention toward teachers and teaching comes with a mandate for rigor, for careful and thorough evaluation, and for the capacity to discriminate knowledge that can generate verifiable solutions at state- and district-level scale from opinions and impressions. Yet despite the acknowledged importance of this need, we could not find a single empirical paper that one would think is essential to evidence-based decisions and policies in the production of effective teaching that was not based on student achievement and market forces.

In a recent controlled evaluation, we examined the extent to which a system of professional development supports that was organized around and focused on CLASS-based measurement and description of teacher-child interactions, using video exemplars of high-quality interaction as well as individualized consultation and coaching (MyTeachingPartner, 2004; Pianta et al., in press) was effective for producing changes in teacher-child interactions and child outcomes. The initial MTP trial tested whether it was possible to provide professional development experiences for early educators that enhanced their interactions with children.

Two models of training were implemented for two years and compared in this longitudinal study: 1) Web Only, in which teachers received workshop training in the curriculum and had access to the video exemplars, and 2) Consultancy, in which teachers received the same workshop training as the other condition but also received the regular, web-mediated consultation. Every two weeks, teachers videotaped their implementation of an instructional activity and shared this footage with the research team. In the consultancy condition, web-mediated interactions between a teacher and consultant then focused on: a) observing the video footage and identifying a teacher’s behaviors with students and their effects; b) problem-
solving to identify and implement alternative approaches as needed and receiving feedback on such attempts; and, c) establishing a nonjudgmental and nonevaluative supportive relationship with a knowledgeable individual (Hadden and Pianta, 2006).

We compared improvement in specific dimensions of teachers’ observed interactions across two intervention conditions (Pianta et al., in press), access to Web Only exemplars or Consultancy. Hierarchical Linear Modeling (HLM) estimated teachers’ growth trajectories for each of the nine CLASS dimensions of teacher-child interactions over the course of the school year. In terms of the main effects of intervention condition on improvement in the quality of teacher-child interactions, all associations were in the expected direction. Teachers exposed to Consultancy supports, in contrast to the Web Only, showed statistically significantly greater improvements in aspects of interaction that involved reading and responding to students’ cues (Teacher Sensitivity), using a variety of formats to actively engage children in instruction (Instructional Learning Formats), and intentionally stimulating language development (Language Modeling).

We also tested whether either of the intervention conditions was more effective in classrooms where a majority of children were from high poverty families. Results indicated that within the highest poverty classrooms, participating in the Consultancy condition was associated with positive changes in Teacher Sensitivity and Instructional Learning Formats while teachers participating in the Web Only condition declined over the year in the quality of their interactions on these dimensions. Thus, it appears that the level or intensity of supports a teacher might need to be successful in the classroom depends in part on the demand characteristics of the classroom itself.

Following evidence that the MTP intervention can alter teacher-child interactions, the next logical question is whether this change in classroom context has any echo in children’s social, academic, and language skills. We used two analytic approaches to address this question: experimental intent-to-treat and quasi-experimental treatment-on-the-treated. Taking full advantage of the RCT study design and two years of intervention implementation, we first evaluated preschoolers’ language and literacy outcomes across the three study conditions; these conditions include the Consultancy and Web Only groups described above, as well as a third control group in which teachers only received curriculum lesson plans (referred to as Materials). Children’s assessed pre-kindergarten language and literacy skills were greater for children in classrooms in which teachers received individual consultation (Consultancy), with an effect size of $d = .21$, compared with the Web Only and Materials conditions (Fan, Pianta, Justice, Hamre, Downer, and Mashburn, 2008).

The quasi-experimental, treatment-on-the-treated analysis exclusively examined teachers in the Consultancy and Web Only conditions (teachers in the Materials condition did not have access to similar resources; Mashburn et al., 2008). During the intervention, these teachers varied in their use of three MTP web-based resources: language/literacy activities, a website that features video exemplars of high-quality classroom interactions, and expert consultation. We therefore examined the associations between teachers’ exposure to these resources and children’s development of language/literacy skills during pre-kindergarten. Controlling for relevant covariates (child, teacher, and classroom characteristics), children showed
significant gains in directly assessed receptive language skills when their teachers received Consultancy support compared with those who were in classrooms of teachers who received no such support. As expected, findings indicated a stronger effect of Consultancy support on children's directly assessed print awareness and teacher-reported language/literacy skills in year two compared with year one.

In summary, there is evidence from our work and others’ (see Jones et al., 2006) that teacher-child interactions can be improved when interactions are the explicit focus of professional development, and that children make significantly greater achievement gains in language and literacy skills when their teacher receives support focused on their interactions with children. In short, basing professional development explicitly on a validated, observational assessment of teachers’ interactive behaviors in the classroom produces improvements in behavior but in children’s learning.

IMPLICATIONS

There is now a reasonable body of evidence, emanating from our work and that of others (see Gordon et al., 2008; McCaslin et al., 2006; Jones et al., 2006), that teachers’ performance in classrooms, in terms of their behavioral interactions with students, can be observed using standardized protocols, analyzed systematically with regard to various sources of error, and in turn prove to be valid for predicting student learning gains. They can also be changed (improved) as a function of specific and aligned supports provided to teachers, and that exposure to such supports is predictive of greater student learning gains. Although modest, these effects are robust and consistent across investigator groups, samples of teachers, and samples of students that vary by grade and socioeconomic and geographic background. Below, we briefly summarize the implications of this work for policy.

We have argued that a major advantage of observational assessments of teachers for leveraging improvements in educational outcomes is that they can be directly related to the investigation and experimentation of specific interventions aimed at improving teaching. For this reason, these methods have considerable promise. Yet measurement challenges, some of which are noted above pertaining to psychometric issues, are not inconsequential. In addition to those challenges described earlier, observational assessments require technical supports that enhance efficiency and lower costs when used at scale. The questions related to psychometrics, efficiency, and costs compel attention and rigorous study, yet the investment in research related to assessments of such “inputs” pales in comparison to research investments in outcomes, specifically standardized tests. Nonetheless, recent Institute of Education Sciences requests for applications do include research on teacher effectiveness and specific topical areas on teacher quality; however, the assessment and measurement goals in the IES framework do not include research on assessment of teachers’ performance in classrooms.

It seems important, at least to us, that investment in measurement studies, cost-efficiency studies, investigations of the determinants and regulators of the quality of teacher-child interactions, and the value of teacher preparation programs for improving such interactions, could be key areas for research and development support. We can envision studies that would identify early predictors of teacher competence, effective supports that improve teaching, virtual reality environments
that accelerate teacher development, and networks of teacher preparation programs studying the natural history and course of teachers’ competence in these performance domains. Although this work could indeed be informed by value-added metrics of teacher quality and effectiveness, we suspect the route to eventual useful interventions and tools will be quicker and perhaps more efficient if focused on teacher-child interactions.

We recognize here the limitations of the work we have described above—measurement issues, small effect sizes, and logistic challenges that impede efficiency and scalability. Yet despite these challenges, the consistency and nature of the results suggests several implications for education policy.

In the realm of accountability, teacher quality and teacher effectiveness are critical “inputs” that counterbalance the focus on student achievement outcomes and potentially address value-laden issues such as equality of opportunity. Given the rather meager results related to teacher characteristics such as education, training, and experience as they relate to outcomes, the results presented herein with regard to observational metrics merit attention as a complementary feature of a comprehensive system.

With regard to production of effective teaching, using teacher preparation programs, credentialing and licensing systems, observational metrics have potential to provide anchors in actual performance that could be used to drive these systems toward higher levels of impact on student performance (providing the observation system is validated). Furthermore, if these systems used observations of performance as an outcome or competence metric, it is highly likely that collateral professional development supports would be developed that were oriented to producing performance at levels specified in licensing and credentialing systems. Finally, performance-based metrics anchored in observations in classrooms may also be compatible with market-based approaches to creating performance incentives, such as merit pay structures.

In sum, we argue that standardized observational approaches to measuring teacher performance represent a credible complement to the current focus on teacher credentials and degrees on the one hand, and value-added metrics of student performance, on the other. Furthermore, observational approaches may link more easily to systems of producing teaching that, in the long run and despite costs and logistic challenges, represent an alternative that has greater long-range benefits for building capacity and quality.
REFERENCES


Using Standardized Observation to Conceptualize, Measure, and Improve Classroom Processes • Robert C. Pianta and Bridget K. Hamre


Using Standardized Observation to Conceptualize, Measure, and Improve Classroom Processes • Robert C. Pianta and Bridget K. Hamre


Weinstein, R. S. (2008). Schools that actualize high expectations for all youth theory for setting change and setting creating. In M. Shinn and H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 81-101). New York: Oxford University Press.
