The Synergy of Class Size Reduction and Classroom Quality

Elizabeth Graue
Erica Rauscher
Melissa Sherfinski
University of Wisconsin

The Elementary School Journal
Volume 110, Number 2

0013-5984/2009/11002-0004$10.00

Abstract

A contextual approach to understanding class size reduction includes attention to both educational inputs and processes. Based on our study of a class size reduction program in Wisconsin we explore the following question: How do class size reduction and classroom quality interact to produce learning opportunities in early elementary classrooms? To address this question, we analyze data from 3 years of fieldwork in 27 classrooms in 9 schools implementing a class size reduction reform. Data generation included multiple ethnographic observations and interviews with teachers and principals, administration of the Classroom Assessment Scoring System, document and artifact collection, and analyses of school-level standardized test scores. We present multiple vignettes to illustrate that class size reduction provides opportunities that can be activated by organizing and implementing high quality classroom practices. We argue that high quality classrooms combined with class size reduction contexts create a synergy for learning.

During a period of economic prosperity, it is easier to imagine and try out new policies. That was the case in the 1990s when governments found discretionary funds to invest in education. Designed to promote student achievement, class size reduction (CSR) programs were among the most commonly implemented reforms. Following the success of the well-documented Student Teacher Achievement Ratio (STAR) experiment in Tennessee, class size reduction seemed to promise an expensive but easy to understand strategy to enhance learning. The federal government and 40 states implemented different forms of class size reduction ranging from programs targeted to particular groups to universal programs mandated for all schools (Graue & Rauscher, 2009). At the same time, critics voiced concerns about the investment in class size reduction programs, arguing that other foci were better proven and more cost effective investments. One suggested alternative was to focus on teacher quality. For those who advocate investments in quality, it makes more sense to invest in the skills a teacher brings to the classroom than to reduce the number of students in the room.

In the education policy version of "tastes great—less filling," class size reduction and teacher quality have been bandied about in the research and policy communities while mountains of evidence have been accumulating. There have been design arguments (What measures of teacher quality are most clearly related to student outcomes?), implementation arguments (How small a group constitutes a class size reduction?), and interpretation arguments (Are pupil-teacher ratio and class size interchangeable ideas?). As research and policy chased one another, it became clear that the questions did not have simple answers.

Instead of positioning class size and teacher quality as competing alternatives, would we learn more if we considered class size and quality concurrently? Could we enhance our knowledge of instructional practices and resource allocation if we explored the potential synergy created by these two elements? Questions like these make sense in a context where researchers are finding that focusing on discrete inputs, like class size or teacher qualifications, has limited utility because learning occurs in the interactions between teachers and students (LaParo, Pianta, & Stuhlman, 2004) with diverse temperaments, dispositions, and expectations. Simple inputs into quality (e.g., such as increasing teacher requirements) have not yielded the outcomes their supporters touted. Focusing inputs into education, removed from their local contexts, is a limited strategy when the goal is to improve quality (Englehart, 2006).

In this paper we explore how class size is related to classroom quality and argue that we must understand both elements if we are to make the most of either education reform. We bring together qualitative data on classroom interactions and processes, standardized measures of classroom quality, and school-level student achievement data to build the case that learning opportunities are created at the intersection of class size reduction logic and classroom practice. To make the case, we analyze data from Wisconsin's Student Achievement Guarantee in Education (SAGE) program, a multi-faceted program that reduced class sizes to a 15:1 ratio in close to 500 schools. Building on research that examined class size reduction mechanisms (Blatchford, 2003a; 2003b; Finn, Pannozzo, & Achilles, 2003; National Institute for Child Health and Development [NICHD], 2004) and the relationship between classroom quality and student outcomes (Rimm-Kaufman, La Paro, Downer, & Pianta, 2005) we examined the following question: *How do class size reduction and classroom quality interact to produce learning opportunities?*

Literature Review

Understanding Quality

Quality in education is difficult to define, codify, and measure. Is quality a function of structural variables, such as teacher qualifications, pupil-teacher ratio, or number of minutes spent on literacy activities, or is it a function of process, such as the types of interactions teachers have with their students, or the nature of activities during 120 minutes of literacy instruction (Stuhlman & Pianta, 2009)? In this section, we turn our attention to understanding structural inputs and process indicators of classroom quality and how these variables relate to research on class size reduction.

Some researchers measure classroom quality through analyses of structural variables. Viewed as "characteristic[s] of the environment that [are] independent of human interaction" (Cassidy et al., 2005, p. 511), structural variables can be measured, quantified, and administratively controlled and regulated. Other researchers examine teacher quality by focusing on teacher attributes, dispositions, skills or practices to explain student success (Kennedy, 2008). Still others focus on student-teacher interaction, believing that "interactions between children and teachers are a primary mechanism through which classroom experiences affect development" (LaParo et al., 2004, p. 412). Studies in this vein focus on quality as generated through social and academic interaction between teacher and child (Hamre, Pianta, Mashburn, & Downer, 2007). It is the most complete conceptualization of quality, and therefore we employ it in this paper.

Process quality has been linked to student behaviors and outcomes (Howes et al., 2008; LaParo et al., 2004; NICHD, 2004; Pianta, LaParo, Payne, Cox, & Bradley, 2002; Rimm-Kaufman et al., 2005; Wilson, Pianta, & Stuhlman, 2007). Instructional and emotional supports have been shown to be particularly important for students at risk of school failure. When high risk students are placed in classrooms with moderate to high levels of instructional and emotional support, they keep pace with low risk peers. Conversely, the achievement gap widened for high-risk students placed in low quality classrooms (Hamre & Pianta, 2005).

In recent years, researchers have connected structure and process through multilevel studies of classrooms. The NICHD Early Child Care Research Network (2005) studied third grade classrooms using a multilevel approach, including teacher behavior and activities, ratings of efforts to improve students' outcomes and teacher characteristics, such as salary and education. Structural inputs (e.g., class size, child-teacher ratio, and teachers' years of experience, in-service training, and salary) explained only about 1% of the variance in classroom quality. Most worrisome was Stuhlman and Pianta's (2009) research based on a large-scale typology of emotional and instructional quality in first grade classrooms, which concluded that (a) instructional support was quite low across classrooms and (b) lower quality classrooms were most likely to include higher need students.

Taken together, this work suggests the need for a comprehensive approach to quality, incorporating analyses that simultaneously attend to structural inputs as the resources that shape teaching and learning and the interactions in the classroom to support learning. Class size might be seen as one of the resources that shape teaching and learning.

Class Size Reduction

Class size reforms are an easy sell because they appeal to parents, teachers, and politicians for many reasons. Popular understanding of class size research asserts that careful implementation of small classes in the primary grades yields positive outcomes for students. Small class effects are typically positive across student characteristics and are greater for students traditionally seen to be at risk (e.g., Biddle & Berliner, 2002; Blatchford, 2003b; Ehrenberg, Brewer, Gamoran, & Willms, 2001; Finn & Achilles, 1990; Finn et al., 2003; Grissmer, 1999; Nye, Hedges, & Konstantopoulos, 2004; Pate-Bain, Achilles, McKenna, & Zaharias, 1992; Smith, Molnar, & Zahorik, 2003). While these patterns are widely cited, there are a number of definitional issues that plague research and policy in this area.

Pupil teacher ratio (PTR) research explores the number of salaried staff serving a group of students without particular attention to teaching practice. The number of students per teacher is treated as a structural input that potentially affects student achievement. Class size (CS) research explores the relationship between teacher practice and the number of students in naturally occurring class sizes. This work often compares relatively larger and smaller classes on variables like grouping strategies or instructional approaches. Finally, CSR focuses on programs that systematically lower the number of students in a class below a particular threshold. It is the intentionality in CSR, a belief that there is a specific value and method in smaller groups that has been the foundation of CSR research (Graue & Rauscher, 2009).

These three approaches are loosely related to three generations of scholarship on the topic (Blatchford, 2003b; Graue, Hatch, Rao, & Oen, 2007). The first generation sought to identify linear relations between class size and student outcomes. This work was designed to answer the question, "Does class size reduction work?" by linking class size with student outcomes. Methodologically, this was deskwork conducted with existing databases that included staffing counts and student outcomes. The input-output model allowed for the comparison of different ratios (including the use of aides in the classroom) and other inputs to education (e.g., teacher qualifications) through cost-benefit analyses.

The second generation of research widened the scope of inquiry by linking class size, teacher action, and student achievement. Focusing on "How does class size affect teacher practice?" this approach contrasted teachers observed with both larger and smaller groups or worked to link specific teacher strategies to student achievement levels. For example, Zahorik, Halbach, Ehrle, and Molnar (2003) examined the practices used by more and less effective teachers and produced a typology of distinct approaches used by each group.

We see our work falling within a third generation of research that widens the scope of inquiry by situating student outcomes at the nexus of teacher action, school culture, and the particular approach taken to instruction within a class size context (Ball & Forzani, 2007; Blatchford, 2003b; Englehart, 2007; Graue et al., 2007). A third generation research question places class size reduction in context:: How does class size reduction interact with other elements of a local context to support teaching and learning? We have been heavily influenced by the work of Peter Blatchford and colleagues (2003b) who describe a contextual approach for understanding the dynamics of class size:

> Our results suggest that it is not just down to the teacher. In contrast to a direct model, it is not entirely the teacher's responsibility; contextual factors cannot be ignored. Teachers will vary in their effectiveness, but the size of the class and the size of the groups in the

> class necessarily affect what a teacher has to deal with, and can present her with choices and the need for compromises. Class size is therefore one environmental contextual factor that will influence teachers and pupils in a number of ways. (p. 160)

A commitment to context is reflected in our assumptions that (a) school culture shapes what teachers can do and (b) administrative as well as instructional decision making contributes to the quality of interactions within classrooms. Teacher practice exists in particular contexts. Professional resources are variably helpful. School culture can enable quality or it can inhibit the work of individual teachers (Graue, et al., 2007).

Recently researchers like Ball and Forzani (2007) have called for the study of the mechanisms of class size reduction:

> Questions about why class size matters to student learning are squarely instructional. They concern what happens inside classrooms when teachers help a smaller number of students negotiate new content. Researchers might consider, for example, the types of instructional activities that teachers are able to plan for and enact when working with only a relatively small number of students or the extent to which smaller class size allows for more productive interactions among students. (p. 532)

Finn et al. (2003) conclude their research review by citing a need for more research on the factors that promote student engagement, the differential impact of class size and other structural features, the role of contextual variables as they affect student engagement, and the direct study of specific mechanisms postulated as influential in class size reduction contexts. Konstantopoulos (2008) recognized this in his recent reanalysis of the Tennessee project STAR data. Asserting interactions between the treatments and individual characteristics, Konstantopoulos pointed to the limits of this work:

> The mechanism is still not clearly defined. Unfortunately data about practices in different types of classrooms are not available. Such detailed observational data could have unveiled the mechanisms of small class size effects via information about instructional processes and interactions between students and teachers. (p. 290)

Class size is a structural input because it is both quantifiable and controllable. However, class size reduction is more than structural. It is about change. Implicit in the theories of <u>why</u> class size reduction is a strategy for improving achievement, is the idea that smaller classes create opportunities that alter fundamental qualities of classroom interactions. Understanding quality in class size reduction contexts is about more than how many or how much; it is also about understanding the processes and interactions of teaching and learning. Focusing on processes and interactions makes class size reduction more complicated and more powerful—it recognizes that the number of people in a classroom is important if it changes instructional interactions. We describe how we approached the task of understanding these processes and interactions in the next section.

Method

This paper comes out of a multiyear, multi-method evaluation of Wisconsin's SAGE program. SAGE is a state-supported class size reduction program that provides $2,250 per low-income child to limit class sizes in almost 500 Wisconsin schools. SAGE is a multi-faceted reform that includes mandates to: (a) reduce class size to 15:1 in grades K-3, (b) provide rigorous curricula, (c) strengthen the links between home and school through keeping the school building open for

extended hours and connecting families with community resources, and (d) enhance teacher professional development and evaluation. The design of the reform recognizes that improving student achievement is a complex challenge that requires equally complex interventions, particularly in communities troubled by racism and poverty (Molnar & Zmrazek, 1994). Originally posed to address the needs of urban schools, SAGE is now open to all Wisconsin schools.

Since 2004 our team has been following a sample of 9 SAGE schools in 6 districts. The sample was purposefully chosen through comparable case selection (Schensul, Schensul, & LeCompte, 1999) to include schools representing a range of economic resources, urban, rural, semi-urban locations, and student achievement. It also represents a range of SAGE implementation strategies, including diverse strategies for reducing class size, providing professional development, connecting home and school, and selecting curricula. Table 1 provides a description of the 9 schools in our sample. The variability across multiple dimensions might be seen as a threat to validity if we were doing a treatment-control study—it would have muddied the causal links that might be asserted. In this case, the variability was a strength of the research design, providing an opportunity to see patterns across the diversity of communities that participated in the SAGE program.

TABLE 1. 2006-2007 SAGE School Sample Characteristics[1]

School	**Farmington**	**Bethany**	**Calloway**	**Earhart**	**Gallows**	**McMahon**	**Montford**	**Wellstone Blvd.**	**West Canton**
Geography	Rural	Urban	Urban	Semi-urban	Urban	Semi-urban	Rural	Urban	Rural
District	Farmington	Mallard	Mallard	Maxwell	Mallard	Bellamy	Walton River	Mallard	West Canton
Enrollment	442	487	300	242	599	233	500	337	306
% Black	0.9	80.3	13.3	31	64.9	31.8	3.8	78.6	1.3
% Hispanic	5.9	1.8	44	14	19	21.9	2.2	6.8	1.6
% White	91	11.1	35	33.5	15	43.8	83.2	6.5	96.4
% Asian	0.5	6.4	5	20.2	0.8	2.1	8.8	6.5	0.7
% ELL	0.2	1.4	1.9	33.9	1.2	13.7	8.4	12.5	n/a
% Students with disabilities	20.4	16	13	8.7	39.6	25.8	16.2	13.1	13.7
% FRPL	58.4	84	78.7	66.1	88.8	68.2	64.4	95.3	37.6
Wisconsin 4th Grade Knowledge & Concepts Test 2006 Percent Proficient & Advanced - Reading									
Mean	83%	82%	85%	82%	53%	79%	75%	66%	95%
Wisconsin 4th Grade Knowledge & Concepts Test 2006 Percent Proficient & Advanced - Math									
Mean	72%	59%	67%	64%	39%	46%	77%	40%	82%
Performance Relative to Expectations[2]									
Test score units	1.24	3.15	5.93	3.46	-5.24	4.82	-1.04	-8.65	1.62

1. All names are pseudonyms.

2. This school performance estimate combines three years of data for each school to provide a more stable and reliable estimate of the expected percent of students proficient or above in grade 3 reading, controlling for student characteristics and average teacher experience and training. Schools with a performance estimate of around zero are performing at expectations, given the population of students who took the tests and the teacher characteristics. Schools with negative performance estimates are doing less well than expected, and schools with positive estimates are doing better than expected

During the 2006-2007 school year our research focused on what we thought of as "best practice." We selected 3 schools for in-depth study that appeared to represent high levels of implementation of SAGE pillars and higher levels of student achievement. Through analysis of prior years' test score data and previous fieldwork, we determined that Calloway (urban), Earhart (semi-urban), and Montford (rural) had recently improved student achievement and seemed to have in place reforms that were changing school culture. In these 3 schools, we returned to a kindergarten, a first grade, and either a second or third grade classroom from our initial sample.

We completed seven half-day observations of instructional practice in each classroom. Our observation protocols were semi-structured, generating field notes that included physical descriptions of space, running accounts of time, counts of students and their activities (monitoring students' entries and exits), grouping, roles, and the actions of adults. We detailed group size and staffing, use of teaching teams, instructional practices, curriculum, and assessment. During observations we sat quietly, taking notes with laptops. We were free to move about the room so that we would have access to both large and small group instruction. We attempted to take verbatim notes of talk and action. The observations were supplemented by photographs of instructional materials and included videotaping[3] of one-half day of instruction. In addition, we observed professional development activities including collaborative teacher planning and instructional sessions. We also observed activities that connected home and school, including family literacy nights and school carnivals. Following each observation, we revised the raw field notes, which were often done in typed shorthand, to a form that could be read and understood by other members of the research team. This revision typically took two to three times as long as the observation period and generated field notes that ranged from 15-25 single spaced pages. Across these various observations, we amassed a minimum of 25 hours per classroom for a total of 225 hours of in-depth qualitative observations.

In addition to observations, we completed multiple interviews with school staff, including two with each principal and three with each teacher. For many this was the third year of interviews for the project and they had a level of comfort with the process that made conversations rich in content. The interviews were semi-structured and focused on the school and classroom practices related to SAGE's four pillars of practice. They also worked to link contextual issues in the district and the school to the specific practice in classrooms. These interviews ranged in length from 40 to 120 minutes and were audio taped and transcribed for analysis. We collected relevant documents and artifacts, including worksheets, curriculum guides, and school and district reports. We also collected school-level test score and demographic data from Wisconsin's Information Network for Successful Schools (WINSS), a publicly available database.

A member of our team conducted standardized observations to rate classroom quality using the Classroom Assessment Scoring System (CLASS). Each classroom was observed for a minimum of 4 cycles amounting to at least 2 hours of documentation. The instrument is described more fully below.

We also collected limited data at the remaining 6 schools from the initial 9-school sample. Visiting 3 classrooms (one kindergarten classroom, one first grade classroom, and either a second or a third grade classroom), we: (a) conducted one interview with each school's

principal and each teacher observed, (b) observed using CLASS, and (c) collected school-level test score and demographic data from WINSS.

Classroom Assessment Scoring System

CLASS (see Pianta, LaParo, & Hamre, 2008) provides a common metric for understanding classroom quality, but moves beyond other rating systems by providing a distinct lens on how classroom interactions function alongside educational resources. We chose CLASS because of its (a) strong empirical record, with use in over 3,000 classrooms pre-K through grade three, (b) its focus on domains identified as important within the literature on class size reduction, and (c) its strong psychometric properties. The domains, dimensions, and behavioral markers employed by CLASS are described in Figure 1.

Emotional Support	Classroom Organization	Instructional Support
Positive Climate The enjoyment and emotional connection that teachers have with students and the nature of peer interactions *relationships *positive affect *positive communication *respect	*Behavior Management* How well teachers monitor, prevent, and redirect behaviors *clear behavioral expectations *proactive *redirection of misbehavior *student behavior	*Concept Development* Measures how teachers promote higher order thinking and problem solving going beyond fact and recall activities with children *analysis and reasoning *creating *integration *connections to real world
Negative Climate Reflects negativity such as anger, hostility, or aggression expressed by teachers and/or students in the classroom *negative affects *punitive control *sarcasm/disrepect *severe negativity	*Productivity* Considers how effectively teachers manage time and create classroom routines that maximize learning time in the classroom *maximizing learning time *routines *transitions *preparation	*Quality of Feedback* Considers how teachers extend student learning through their responses and participation in activities *scaffolding *feedback loops *prompting thought processes *providing information *encouragement and affirmation
Regard for Student Perspectives Captures the degree to which teachers' interactions with students and classroom activities place an emphasis on students' interests, motivations, and points of view *flexibility and student focus *support for autonomy and leadership *student expression *restriction of movement	*Instructional Learning Formats* Focuses on how teachers engage students in activities and facilitate activities so that student learning is maximized *effective facilitation *variety of modalities *student interest *clarity of learning objectives	*Language Modeling* Reflects the extent to which teachers facilitate and encourage student language *frequent conversations *open-ended questions *repetation and extension *self- and parallel talk *advanced language
Teacher Sensitivity Reflects teachers' responsivity to individual students' academic and emotional needs *awareness *responsiveness *addresses problems *student comfort		

The CLASS manual (Pianta et al., 2008) provides a variety of evidence of its reliability and validity. We summarize a sample of the evidence here. Table 2 presents descriptive information of CLASS use in a number of major studies and with the SAGE sample. The ratings of CLASS dimensions and domains are higher in SAGE classrooms than in the empirical sample, with few classrooms rated in the low range. Given our research design, we cannot directly link CLASS ratings to SAGE's influence because we do not have data from comparable non-SAGE classrooms.

TABLE 2. Mean Class/Domain Scores for Sample SAGE Schools and CLASS Empirical Sample[1]

	SAGE Sample n = 27		CLASS K n = 730		CLASS 1-5 (Responsive Classroom) n = 88		CLASS Grade 3 n = 82	
CLASS Dimension[2]	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Positive Climate	5.49	1.2	5.14	NR	4.91	.93	4.44	1.17
Negative Climate**	1.48	.56	1.55	.65	1.35	.65	2.22	1.16
Teacher Sensitivity	5.03	1.34	4.64	.86	4.74	1.01	4.60	1.10
Regard for Student Perspectives	4.67	1.22	NR		NR		4.28	1.08
Behavior Management	5.36	1.17	5.18	.79	5.14	.95	4.98	1.29
Productivity	5.65	1.09	4.67	.73	4.98	1.00	4.69	1.14
Instructional Learning Formats	4.95	1.12	4.11	.84	4.23	.73	4.21	1.22
Concept Development	4.35	1.13	2.11	.74	3.82	1.01	3.84	1.32
Quality of Feedback	4.35	1.12	1.84	.64	4.77	1.03	3.54	1.31
Language Modeling	4.32	1.04	NR		NR			
CLASS Domain								
Emotional Support	5.43		5.41[4]		5.92		4.78	
Classroom Organization	5.32		4.65		5.06		4.63	
Instructional Support	4.34		1.98		3.26		3.69	
Average CLASS SCORE	5.10		4.27		4.74		4.50	

Confirmatory factor analyses explored the internal consistency of the three main domains measured by CLASS. Reported separately for five large studies, alphas were in the moderate to high range (.76-.95) across 3 domains. Reliability was assessed across cycles, with ratings in the moderate to high range (r = .68-.97) comparing first cycle to the total score. With 2 cycles, correlations with the final score rose to above .90. Measures of stability indicated that CLASS scores were highly stable across days with correlations ranging from .73 to .85. Instructional Support ratings predicted child language, pre-reading concepts, and applied math skills, while Emotional Support ratings were associated with language skills and teacher-reported behavior problems (Pianta et al., 2008).

While all research team members completed in-depth training and were certified CLASS coders,[5] one team member completed all CLASS observations in this data set. Her classroom visits were scheduled to coincide with typical instructional practice with minimal interruptions for other activities (e.g., recess, school programs, music/art/physical education classes, lunch). In each classroom, the rater observed and coded classroom quality across four thirty-minute cycles of instruction with CLASS.[6] These observations generated brief field notes that the observer used to assign ratings of 1-7 on each of the CLASS dimensions, referring to detailed rubrics in the CLASS manual. Ratings are categorized at three levels: 1-2 = low, 3-5 = midrange, 6-7 = high. Average scores were tabulated across four cycles and CLASS dimension and domain scores were calculated by both classroom and school.

For each of the three case study schools, our field notes from each classroom were used as a crosscheck for CLASS ratings. We compared the short notes generated during a CLASS observation with the field notes generated in the case study to ascertain that the CLASS observation was of typical practice. We then discussed the rating to compare our understandings of the site and the evaluation of classroom quality. In each case the rating of the assigned class rater came within 1 point of the ratings suggested by the outside rater (this is the criterion established for reliability by CLASS). It could be said of this strategy that the team of three researchers conducted a qualitative form of reliability check that compared CLASS ratings to a large set of field notes generated across time.

It is important to note that CLASS is not a measure of teacher quality but rather of classroom quality. This is relevant because it recognizes that teacher skill and knowledge are mediated by structural factors like physical space, schedules, supplementary staffing, and the availability of materials. Classroom quality, as experienced by individual students, is funneled through all of these resources. For example, during rating periods when there are multiple adults in a classroom, CLASS raters focus on the average experience of students in the classroom, using the same indicators as would be used with a single teacher.

To illustrate how CLASS works, we present an example of whole-class instruction in a highly rated SAGE first grade classroom. Observed indicators of the three CLASS quality domains (Emotional Support, Classroom Organization, and Instructional Support) are provided alongside field notes to show how classroom quality is analyzed using CLASS. Note that this is not an exhaustive CLASS analysis. Our goal is to offer a concrete example of how CLASS constructs and rates classroom quality. In the vignette analysis seen in Figure 2, the left column

details CLASS criteria and corresponding ratings reflective of the classroom are described on the right[7]:

CLASS Ratings with Representative Scores

Classroom Organization/Instructional Learning Formats/Productivity—The teacher actively and authentically facilitates student engagement by sharing a letter from a former classmate and effectively using a bit of dramatic flair. Later, hands "shoot up" signaling engagement. Uses a routine activity for instructional purposes (High Range, 6-7)

Instructional Support/Concept Development & Language Modeling—Concepts are related to students' lives through the letter. Ms. H.'s open-ended questions spark analysis. (High, 6-7)

Emotional Support/Positive Climate & Teacher Sensitivity—Ms. H. is aware of Emily's need for time and is warmly supportive. (High, 6-7)

Support/Quality of Feedback—The teacher rephrases and extends Emily's response. (High, 6-7)

Instructional Support/Concept Development—Ms. H. links literacy concepts to the chart generated through prior lessons. (High, 6-7)

Emotional Support/Regard for Student Perspectives—The class collaborated over time to make the *Ways to Use Words* chart. It serves as a powerful learning tool across lessons and independent work. (High, 6-7)

Emotional Support/Teacher Sensitivity—The teacher describes Wendy's feelings and acknowledges emotions the class may experience regarding a friend's move. (High, 6-7)

Classroom Organization/Productivity & Behavior Management—Ms. H. is well organized with materials for a smooth transition. She never needs to give a behavior cue. (High, 6-7)

Instructional Learning Formats—Materials for reading are highly engaging to the students. (High, 6-7)

Classroom Practice Vignette: Tammy Helman's (Earhart) 1st Grade Morning Meeting

The students gather for a morning meeting and Ms. Helman's read a letter:

"I have a letter I want to share with you and it says on the outside, 'Do not open until 5/27/07.' Today is 5/29 but we weren't here on the weekend for 5/27 so we can open it today. It is to all of us from our friend Wendy." Ms. Helman carefully opens the letter and reads, "Dear Class, I wish I could stay, but I can't, I can't, I can't!" She laughs and makes a sad face. "Your friend, Wendy. P.S. Please cheer up!"

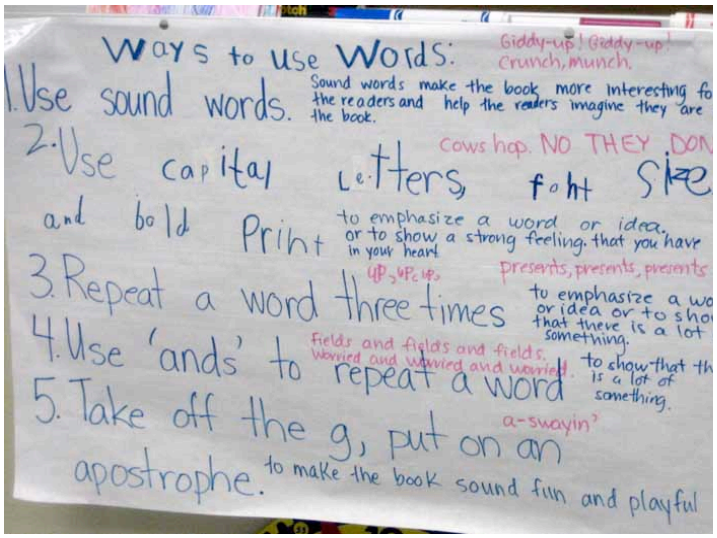
Ms. Helman leans toward the group, and pretends like she's going to cry. "I wonder why she said 'I can't. I can't. I can't' three times like that? What do you think she was trying to tell us?"

Several hands shoot up. Ms. Helman calls on Emily who replies that Wendy was thinking that they wouldn't want her to leave so she wanted everyone to cheer up. Her response meanders a bit and Ms. Helman patiently waits for her to come to the point. Ms. Helman shakes her head yes, suggesting that is why she put the P.S. because they had all been very sad when they discovered she was leaving.

"What about this 'I can't, I can't, I can't'? What was she trying to emphasize?"

She calls on Matthew who says, "Three words."

For those unfamiliar with this classroom, this response wouldn't make much sense. But Ms. Helman turns and points to the poster the class had made about strategies for writing titled *Ways to Use Words*.



> "She repeated it three times. She just can't stay—she doesn't have control over the situation because her family had to move. I'll hang it up here so you can look at this later. It sure is sad when our friends move."
>
> Ms. Helman hangs the letter up on the chalkboard at a height that allows children to read it, then turns to the next item on the agenda—a story about Indian elephants that rob cars of food.

This first grade classroom exemplifies high quality by demonstrating CLASS dimensions enacted at a high level and frequency. Classrooms with moderate and low ratings would display lower levels of the dimensions and fewer incidences. While this classroom had consistently high quality interactions, it is not uncommon for a classroom's scores to vary among domains and dimensions. The standardized nature of CLASS allowed our research team to share a common language about quality and to operationalize ideas in our research that for two years had been only tacit.

Analyzing Diverse Data

Analysis and data collection formed a recursive cycle shaped by the constructs that structured the study design while taking up themes that emerged from fieldwork. Supported by the qualitative data analysis program NVivo, the research team coded data sources using a shared set of codes. These codes reflected the assumptions that guided the research as well as concepts that were new to the project. As the year progressed, the research team met to share observations and interviews and suggest emerging themes. We then returned to the field with those themes in mind. Through this process our focus shifted from general to specific and back again, depending on what we were learning. We shared memos (Graue & Walsh, 1998; Maxwell, 1996) that detailed analytical conceptualizations that linked coding categories and illustrated crosscutting themes. We also introduced some of our emerging analyses with our participants, who provided grounded reactions to readings their experiences.

The specific work for this paper began with the examination of the CLASS ratings of quality, looking for patterns within structural elements such as school achievement and class size reduction configurations. We worked between CLASS quality ratings and our case study data to construct classroom quality vignettes (Erickson, 1990; Graue & Walsh, 1998) to illustrate how opportunities afforded by class size reduction were variably realized. Derived from field notes and edited for ease of reading, the vignettes depict representative experiences observed through general fieldwork and illustrated classroom quality rated by CLASS. In 9 of the classrooms, the CLASS rater and fieldworker were there at the same time providing another cross check. Our use of CLASS was complemented by 3 years of data collection in these sites, including extensive field notes, interviews, and artifacts providing a multilayered source for analysis.

This approach had both inductive and deductive elements. From 3 years of data generation in these 9 schools, we had a sense of the tensions among simple structural explanations for the SAGE program's success (Graue et al., 2007; Webb & Meyer, 2004) and links among school and individual teacher practice and student outcomes. We observed that the variation within schools or SAGE configurations was often as great as the variation between schools. It was therefore important to closely examine patterns in administrative practice, teacher beliefs, types of classroom interaction, and school achievement. In contrast, we also worked from the ratings to the case study data, exploring likenesses and differences in high, midrange, and low rated classrooms. We looked for confirming and disconfirming evidence for the patterns

we identified. This was particularly powerful as we had team members who had long-term involvement in each site and the CLASS rater who had visited all sites. This allowed us to work between depth and breadth of knowledge across the team to triangulate the assertions made at each phase of our research.

The construction of this paper required careful consideration of argument staging and data use. Three years of fieldwork yielded massive amounts of information and we had understandings of contexts and relationships that framed our reading of the 2006-2007 data collection. We began by identifying examples of different class size reduction configurations in our sample and we paired those types with different ratings of classroom quality. We chose extreme examples of 15:1 and 30:2 configurations, focusing on high and low rated classrooms and then found examples that illustrated the CLASS ratings. We provided contextual information with the three vignettes chosen to illustrate the interrelationship between classroom practice and the broader context. The vignettes presented were taken from our case study schools but represent practices observed in other schools as well.

Given the mixed methods approach chosen here, we need to be clear about the grounds on which we made inferences. We took a strong interpretive approach, with attention to the local enactment of the SAGE program in a purposefully chosen sample of Wisconsin schools. We do not pretend that this sample was designed to allow statistical inference; therefore our interpretations are applied to neither all SAGE schools nor all class size reduction programs. We rely on the notions of trustworthiness and transferability, where the readers make connections from the case to their own knowledge and experience, evaluated by the degree to which the findings appear to resonate with the reader from their reading of the existing literature and the presentation of the case relative to the literature (Lincoln & Guba, 1985).

Results

Our first cut at the data explored the patterns of CLASS scores within a school to determine if high or low quality classrooms were clustered. This relationship is depicted in Figure 3. The range of quality in schools varied on a CLASS metric from .57 (very similar) to 2.43 (most likely a difference in category). For example, Bethany and McMahon schools displayed fairly similar CLASS scores across classrooms, while others were much more varied within a school.

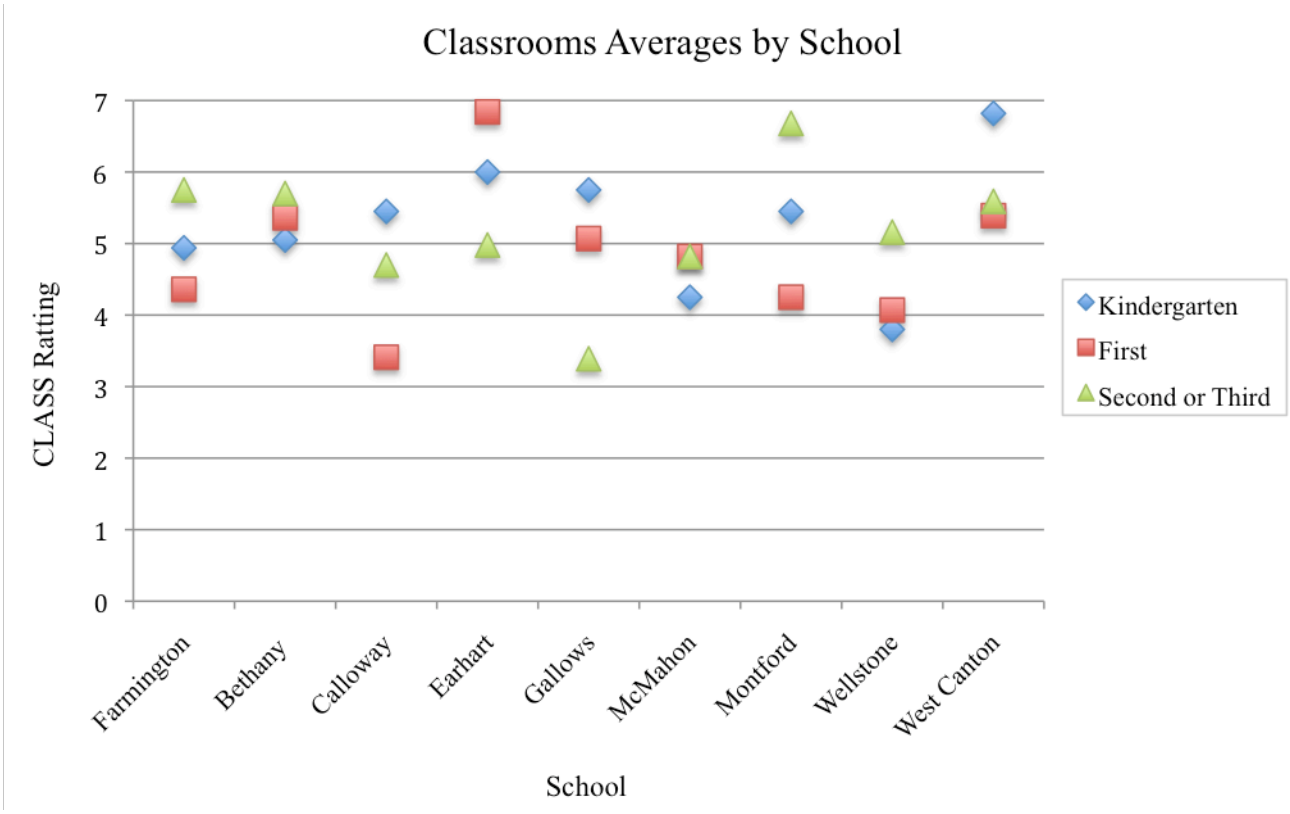


Figure 3 illustrates the within school variation that has challenged us to think between school level and classroom level analysis. Recognizing that variation within schools was sometimes greater than between schools prompted us to examine quality at the classroom level with variable school supports.

CLASS and School Level Achievement Data

We became interested in measures of classroom quality to better understand the affordances created by SAGE, and by extension, other class size reduction initiatives. It was clear that analyses of student performance were incomplete ways to understand class size reduction. However, we also recognize that an analysis of the policy without incorporating all available data could be seen as equally incomplete, so we include its analysis in this discussion.

For each school in our sample, we examined CLASS ratings by school achievement[8] on the Wisconsin Knowledge Concepts Examination (WKCE) state fourth grade reading test, given in the fall of 2006. We contrast school-level achievement data from WINNS in terms of three categories: high (80% or more scoring at the proficient or advanced levels), moderate (70-79% scoring at the proficient or advanced levels), and low (less than 70% scoring at proficient or advanced levels). The results of this analysis are presented in Table 3.

TABLE. 3 CLASS Domains by School Achievement on State Grade Four Reading Test

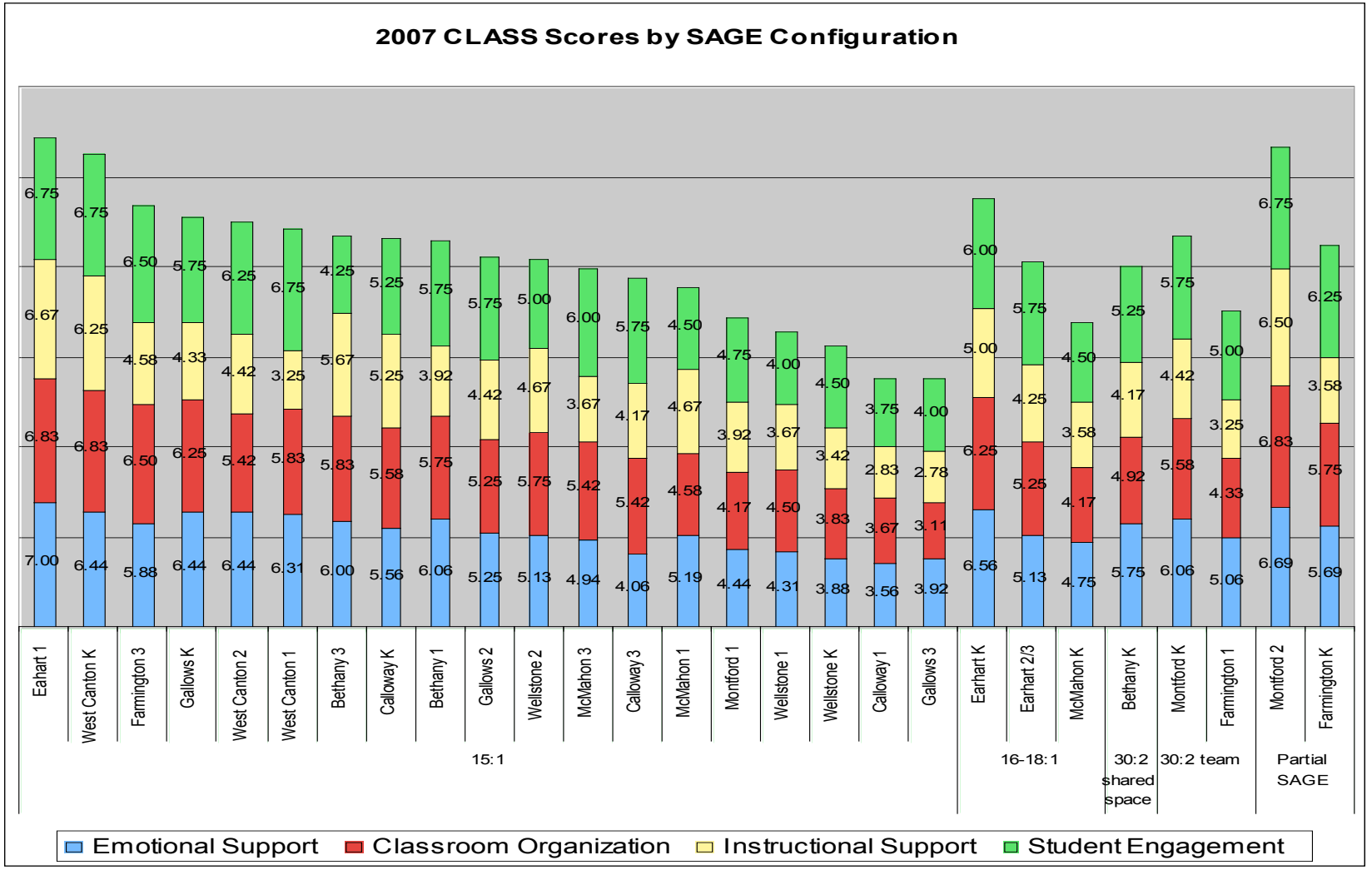
Student achievement on WKCE	Emotional Support	Classroom Organization	Instructional Support	Student Engagement	CLASS Mean
High (n = 5)	5.70	5.61	4.76	5.73	5.38
Moderate (n = 2)	5.34	5.13	4.46	5.38	5.08
Low (n = 2)	4.82	4.78	3.88	4.86	4.58

High = >80% proficient and advanced
Moderate = 70-79% proficient and advanced
Low = <70% proficient and advanced

The ratings follow the expected pattern, with higher CLASS ratings for high performing schools, followed by moderately performing schools, then low performing schools. This is an indication that CLASS, aggregated at the school level, provides a measure of quality that is related to student achievement.[9]

Classroom Quality and SAGE Configurations

SAGE implementation of class size reduction and these implementations changed from year to year. The majority of the SAGE classrooms we studied had one teacher paired with 15 (or fewer) students. Some stretched the limit of 15 students up to 17. Other classrooms used two teachers, either paired in a team teaching situation with 30 students or dividing the room into two separate instructional spaces. Finally, some schools were granted special waivers by the Wisconsin Department of Public Instruction (DPI) to reduce class sizes only for core programming, with smaller groups during math and literacy instruction. This focus on a ratio might be seen as a PTR approach; however, we analyze it as an example of a CSR reform. While PTR research typically calculates a ratio at the school level, our unit of analysis in this study is the classroom. We were interested in how SAGE configurations set a context for classroom processes. Our analyses of quality by configuration are presented in Figure 4.


2007 CLASS Scores by SAGE Configuration
Eahart 1
West Canton K
Farmington 3
Gallows K
West Canton 2
West Canton 1
Bethany 3
Calloway K
Bethany 1
Gallows 2
Wellstone 2
McMahon 3
Calloway 3
McMahon 1
Montford 1
Wellstone 1
Wellstone K
Calloway 1
Gallows 3
Earhart K
Earhart 2/3
McMahon K
Bethany K
Montford K
Farmington 1
Montford 2
Farmington K
15:1
16-18:1
30:2 shared space
30:2 team
Partial SAGE
Emotional Support
Classroom Organization
Instructional Support
Student Engagement

Each bar in Figure 4 represents CLASS scores for individual classrooms labeled by the school name and grade level. On the far left are five classrooms that paired two teachers with large groups of students. To the right are smaller classes with one teacher. The key to interpreting this figure is to look at the big picture illustrated by the length of the bars. For example, the Earhart first grade classroom and Montford second grade classroom had especially high CLASS scores overall. In contrast, the first grade classroom at Calloway and third grade classroom at Gallows had much lower CLASS scores. Paralleling the work on quality nationally (Pianta, Belsky, Houts, Morrison, & NICHD, 2007), this is a story of variation within and across groups. No one configuration has higher CLASS ratings, so we can make no conjectures about the relationship of structural variable of group size and classroom quality. To gain more insight into that relationship, we explored the interviews and observations.

In the section that follows, we contrast quality at the classroom level with comparisons of moderate and high quality. Our analyses are designed to mirror the variation in quality in different class size reduction contexts that would be missed by the typical structural, input-focused analyses. This variation is important because it illustrates the range of quality within classrooms. We emphasize again that our analysis is interpretive and does not assert causal mechanisms.

Profiles of Quality

In this section we present profiles of classroom quality in the context of SAGE-style class size reduction. We first compare structurally similar cases by describing quality in 15:1 classrooms. We then present a team teaching profile representative of average 30:2 quality for our sample. Analysis is attuned to both contextual and CLASS quality factors.

15:1 in practice. The 15:1 first grade classrooms we highlight are from schools that would be seen as quite similar. They are small to mid-sized schools with similar student achievement, both performing better than expected on state tests. We focus closely on quality in two first-grade classrooms—Calloway's Gloria Howard's and Earhart's Tammy Helman's—to illustrate how quality shapes the power of class size reduction. The two cases were chosen by extreme case selection (Schensul et al., 1999), representing high and low ratings of quality in our sample.

Calloway Academy was a midsize urban school with a diverse population. Calloway's principal was Mrs. Collier, a strong leader who organized and monitored all aspects of school life. Mrs. Collier was the "decider" about all things at Calloway and teachers complained that they felt they did not have a voice at the school. In this top down culture, there were few resources to build capacity from the classroom up. This caused problems when teachers struggled. Quality ratings in Mrs. Howard's classroom were among the lowest in our sample, showing a context that was emotionally (ES = 3.56), organizationally (CO = 3.67), and instructionally (IS = 2.83) in need of improvement. She saw her small class of 13 students as full of challenges—behavior problems, second language issues, and too many boys (Howard interview, midyear). Despite advice from colleagues to use strategies like rearranging the environment to manage behavior, by April Mrs. Howard was still struggling.

For literacy instruction Mrs. Howard used the same text with all ability groups, varying the pacing of instruction rather than the content. This is a description of small group instruction with the low reading group while the rest of the class was assigned seatwork:

> Mrs. Howard hands students a small book and asks them to open to Seal Beach on page 17. Mrs. Howard tells them to put their fingers on the title and read slowly, "S-eee-l B-eee-ch" and asks, "What will this story be about?" After they make predictions and read the About the Author section, Mrs. Howard asks Max to start reading. Max begins, "Mom and I . . . (Mrs. Howard: www) walked to the --- sea." (Mrs. Howard: s-ea. That's a period.) The kids know from prior experience that they are supposed to take turns. Vicki reads, as Mrs. Howard helps, "We take st . . . (Mrs. Howard: steep) st (Mrs. Howard: steep steps) that (Mrs. Howard: l-eee) lead to the beach." Mrs. Howard tells Vicki that she did a good job on beach. Edwin is next, "We see . . ." as he pauses, Mrs. Howard says, "Look at the picture Edwin, 'We see seeeels.'" Edwin finishes his turn by repeating the word "seals."
>
> The rising noise level draws Mrs. Howard's attention away from the group. She redirects George, who is peeking from behind the partition that surrounds his desk. When she calls on Nabil, he replies that he is finished with his seatwork (meaning he would like to go to the activity centers). When Mrs. Howard says that she needs to see his work first, Nabil brings his papers to Mrs. Howard. The kids in the reading group keep reading, but when they get stuck on a word they stop and wait for Mrs. Howard to turn her attention back to the group. After looking at his work, Mrs. Howard tells Nabil that he has to redo some of his handwriting paper. Edwin is copying words off the board. Derek is cutting a crayon. Haylee keeps getting glue. Max is crawling around the room combat style. Mrs. Howard notices and barks, "Max, get in that seat!"

This short segment is representative of lower moderate CLASS scores. The room is not emotionally supportive, with few positive verbal or physical interactions and little sensitivity to the children's needs. Classroom organization is ragged. The students experience periods of empty time as they wait for permission or input. Because students have not yet internalized independent work routines, some are off task or interrupt the reading group. In response, behavioral feedback is often terse and Mrs. Howard raises her voice to maintain discipline. The students depend on step-by-step directions. Instructional support in this classroom is quite limited—Mrs. Howard gives them words as they read and feedback focuses on decoding. While there is some discussion of the text and prediction at the outset of the lesson, there is no further development of understanding along the way.

We contrasted this with work at Earhart Elementary, a small school in a semi-urban district. Earhart was led by a principal who worked to include her staff in all decision making. One of the most respected teachers in the school was Tammy Helman who worked with 13 first graders. Introduced earlier in the paper through the letter example in Figure 2, this classroom had uniformly high CLASS ratings (ES = 7.0, CO = 6.83, IS = 6.67). In a highly integrated reading and writing program, students develop dispositions that allow them to regulate their own behavior and find resources in the classroom.

In the following example, we see this in action as Ms. Helman works with Esme to make sense of a text during independent reading time. This example illustrates a positive climate that builds on children's needs and interests, a productive classroom that runs like a well-oiled

machine, and conceptually rich instruction that connects ideas to real world experience through a scaffolded conversation:

> The morning routine begins with students reading on their own as Ms. Helman works through routines like attendance and lunch count.
>
> "Is this 'presented'?" asks Esme. Ms. Helman touches her arm gently and says, "Why don't you go back to your seat and see if that makes sense, I'll come read with you." Scanning the students busily at work, Ms. Helman warmly says, "I'd like to thank the people who are reading. Thank you Nancy, thank you Antoine, thank you Emily."
>
> Ms. Helman squats down next to Esme and says, "Find the first part of this word—it's kind of like bee or me."
>
> Esme crinkles her nose and says, "pre?"
>
> "What about the rest of the word?"
>
> "T—ended . . . pre-t-ended . . . pretended?"
>
> Ms. Helman cocks her head and looks closely at Esme, "Does that make sense? Why don't you read the whole sentence?"
>
> "He pretended to be a jack-in-the-box."
>
> "OK—does that make sense?" Esme nods yes. "If he is pretending to be a jack-in-the-box, what is he doing?" Ms. Helman mimes using a jack-in-the-box and encourages her to keep reading.
>
> "Sometimes they did magic tricks wearing a sw----ir----ling swirling purple cap."
>
> "Look at what she's wearing right here," Ms. Helman says pointing to the picture, "this purple thing. It is called a cape. A swirling purple cape. Do you know of any other characters you have seen who wear a cape? . . . Like a magician maybe?"
>
> "I saw a black one of these on TV once."
>
> Ms. Helman smiles at Esme and turns to Destiny who is juggling her breakfast tray, "Here Destiny, let me help you with that." Esme is reading happily.
>
> Across the room Alison uses red disks to count out the number of adjectives she has included in her book of riddles. Ernie decorates the cover of the book he is writing by drawing a heart around his name twice. Erica glances up and notices that her friend Antoine has just finished his storybook. She whispers, "If you are done with it you have to read another book." Alison is busy editing—ogling her little red counters while whisper-reading sentences to herself and playing with her use of "juicy" words.

This short example shows strong language modeling, conceptual development, and feedback loops between a teacher and student. Ms. Helman begins this 10-minute tutoring session by recognizing students who were working independently then focusing her full attention on Esme. The rest of the class is engaged in work and there are no interruptions. This kind of individualized instruction is possible because Ms. Helman is not the sole source of knowledge in the classroom. Through the year she had worked to create a climate where students can work independently and support each other so no time in the classroom is wasted. In contrast to the first example, in which instruction is fractured by interruptions, this class uses the class size

reduction resource optimally through classroom organization that allows Ms. Helman to focus on one student's needs. Ms. Helman uses a variety of instructional support strategies—questions, thinking prompts, visual cues, repeating and extending responses, and relating the content to Esme's experiences.

From a teacher quality perspective, Ms. Helman was a superb teacher, but exploring classroom quality provides a more comprehensive view. Although we do not provide the full case for the institutional context at Earhart, we can provide a thumbnail sketch to complement our description of this first grade classroom. Ms. Helman's considerable skill was complemented by professional development provided by her district, a school culture of problem solving, adequate classroom space, and a commitment to assessment-guided instruction (Graue et al., 2007; Graue & Rauscher, 2009). The quality in this classroom was enhanced in an institutional context that supported teacher growth and that provided opportunities for intense, appropriate instruction. The resource of class size reduction, provided by SAGE, was part of this quality context.

30:2 team teaching. Team-taught classroom configurations are often used out of administrative necessity, motivated primarily by a lack of school space and rarely seen as a perfect solution. While space precludes a full description of different team teaching situations, we introduce a few key issues from our fieldwork that point to unique issues in 30:2 configurations, using cases from Montford (a kindergarten class) and Farmington (a first grade class) as exemplars.

Montford has a shared leadership culture in which the teachers and principal work to develop programs that are responsive to student needs. SAGE was part of this plan. However, the school building was not quite big enough to accommodate each teacher having his or her own classroom. As a result, several SAGE teachers had developed long-standing partnerships, sharing a classroom, planning, and instruction. According to Principal Mary Durst, the Montford model of SAGE was premised on using all available resources:

> What I always want to know is, are we using the teachers to best meet the needs of the kids? And SAGE doesn't always mean 15 to 1. It might mean 24-to-1 and 1-to-1, you know if there are two people in there. . . . If one-to-one best meets that child's needs and you can still meet the needs of the rest of the 24 you do that too. But what you don't want it to be is 25 to 1 and one sitting in the teacher's lounge.

Instructional practices at Montford were supported by a comprehensive assessment plan facilitated by Dina Monroe. As the school's professional development coordinator and coach, Mrs. Monroe managed data for her colleagues by linking student assessments with strategies for instruction based on the consideration of both goals and students' needs. This approach reflected the school-wide use of Responsive Classroom (see Rimm-Kaufman, 2006), an asset-based approach to teaching and relationship building that emphasizes the social and emotional dimensions of instruction.

The classroom taught by kindergarten team teachers Karen Martin and Nancy Giles was rated within the high and moderate quality ratings across CLASS domains (ES = 6.06, CO = 5.58, IS = 4.42). Working together for a number of years in a school pressed for space, they had developed a strong curricular base emphasizing hands-on learning centers, responsiveness to student interests and needs, and a keen sense of professional respect punctuated with good humor. They scavenged space around the school since their classroom was not big enough for

individual or small group work with 26 boisterous kindergartners. The teachers described their partnership like a marriage—a matter of constant negotiation. Mrs. Giles and Mrs. Martin capitalized on having two teachers in a classroom and their rich knowledge of student lives and needs to enhance instruction. This example illustrates strength in teacher regard for student perspectives, concept development, and diverse learning formats.

> Mrs. Giles is reading to the class and Mrs. Martin is on the floor with 26 kindergartners. "What's 'stagger'?" a child asks after hearing the word in a story. While Mrs. Giles doesn't hear the question, Mrs. Martin does and she repeats it. With dramatic flair, Mrs. Giles then rises from her chair and demonstrates, staggering across the classroom.
>
> The group circles the carpet—all students and 2 teachers are ready for sharing time. Noa suggests they do "zoom" sharing. After Mrs. Martin asks Noa if he means "whip share,"[10] she announces the sharing question: "What is your favorite color popsicle?" The children call out their favorite color as the focus zooms from child to child. Chue says his favorite color popsicle is black. Mrs. Martin asks him if he really means popsicles because she doesn't know of any popsicles that are black. He is silent and she describes a popsicles—Chue seems to know then and changes his answer to blue.
>
> Each teacher then takes thirteen students for Daily News. The two groups engage in identically planned lessons, Mrs. Giles' in the classroom and Mrs. Martin's in a makeshift space under a staircase. In smaller groups, there is much opportunity for dialogue:
>
> Mrs. Martin starts out with the Daily News, sung to the tune of "Popeye the Sailor Man." She spiritedly blows a train conductor's whistle for the "toot toot" part.
>
> <u>It's time for the Daily News (toot toot)</u>
>
> <u>It's time for the Daily News (toot toot)</u>
>
> <u>So sit on your pork chops and we will get started,</u>
>
> <u>It's time for the Daily News!</u>
>
> About to have a student read a message from the whiteboard, Mrs. Martin is interrupted when Gabriel shouts, "Mrs. M! You have to look outside!" Outside of the door is a big red garbage truck. Mrs. Martin looks out the doorway and then asks Gabriel, "What is it?" Gabriel replies, "A garbage truck!" Mrs. Martin makes the most of what could be seen as a distraction. She asks Gabriel, "What letter does it start with?" Just as the hard /g/ sound leaves his tongue, Gabriel's eyes register knowing. He shouts, "G! Like Gabriel!" Mrs. Martin smiles and continues with the Daily News. Gabriel is beaming.

Children's learning is fostered in this classroom when academic content is joined with the experiential content of the children's lives. Children's perspectives are taken seriously and their curiosity presents "teachable moments" that the teachers embrace. This is a lively and positive classroom climate where teachers regularly weave drama, games, language experience, and music into engaging instructional learning formats. English language learners (ELL) and native English speakers are assisted in their language development through practice opportunities. There are many spontaneous conversations in this classroom. Tight quarters are handled

creatively by expanding out to the hallway nook. This kindergarten classroom's quality of organization includes clear behavior management and smooth transitions. This level of quality is possible in a school with shared leadership that builds on the expertise of staff, while building capacity to meet the needs of individual children and families.

Farmington was a rural school much like Montford that experienced similar space challenges. Principal Catherine Paley told us that she chose team teaching for her staff because: "My own personal bias is that for some students, having two teachers in a classroom can really benefit them. I think there is something that two people can give you that one person cannot. In some respects I think there's always somebody there that you can talk to. I also think if you can get teachers who complement each other in their styles and in their delivery that's good for kids."

Several teachers at Farmington looped with their students, taking them from one grade to another. Looping is favored for its relationship building and continuity (Burke, 1996), but in the SAGE context it had some unintended consequences. With staffing that was not absolutely stable and the need to pair teachers at a grade level, teachers often found themselves with a new partner each year. This was the case for Dana Read, a first grade teacher who had looped from kindergarten to first grade. Her classroom seemed to have a revolving door: "We're in a situation where I'm co-teaching with somebody completely different than I've ever co-taught with before. That's been a challenge just because we are polar opposites (laughs) . . . I have taught with a different person every year for the past four years." This lack of continuity had a palpable effect, disrupting the synchrony needed to make the most of teaming and instruction tended to be characterized by fits and starts. While we do not provide a vignette of classroom quality in Mrs. Read and Mrs. Prange's classroom in this article, when we contrast their CLASS profile (ES = 5.06, CO = 4.33, IS = 3.25) with the other 30:2 example, we see a relatively warm emotional context, with moderate classroom organization that has limited instructional support. We suggest that this can be explained in part by the incomplete use of two teachers in the classroom.

Summary. Our profiles of quality show variation on two dimensions. The variation in class size reduction implementation includes settings with both small and large groups and with single and teamed teachers. In addition there is variation in CLASS-assessed quality. Taken together we see that there is not a linear relationship between these two dimensions, that high or moderate/low quality can be seen in either configuration. This highlights two important points. Class size reduction in the SAGE reform is not a single treatment and any attempt to understand program effects must take this into account. In addition, classroom quality is supported by a complex combination of teacher skill, knowledge, disposition, administrative strategies, and physical space. Bringing these two elements together helps us understand their particular synergy in a richer way.

Discussion

Research on teaching and pedagogy has shown a lack of interest in classroom contextual influences. There has been an underlying assumption in many studies of a direct model, with teaching affecting pupils' achievements and learning in a causal way (Blatchford, Kutnick, Baines, & Galton, 2003). Some academics and policymakers have also downplayed the importance of class size differences (Rivkin, Hanushek, & Kain, 2000). Yet teachers do not meet pupils out of context, and class size, or the number of children in the classroom, can be seen as a contextual influence on classroom life, one that plays a

> part in affecting the behavior of teachers and pupils. (Blatchford, Bassett, & Brown, 2005, p. 454)

As class size reduction scholarship has matured, there has been a growing recognition of the importance of taking a contextual or systems approach to understanding CSR reforms. While the number of students in a class can be a starting point for what teachers do, there are other elements that set the stage for the schooling experience. In earlier work we explored alternatives like physical space, staffing patterns, and administrative decision making and their effects on teaching and learning in a class size reduction program (Graue et al., 2007). We recognized that if we were to take seriously Blatchford and colleagues' assertion that context matters in studies of class size reduction, we needed to look for a better way to conceptualize what happens in classrooms. CLASS emerged as the most suitable tool because it helped us attend to multiple dimensions of students' opportunity structures.

Participation in a CSR program is one of many elements that can change schools and classrooms. Our research to date and readings of the literature lead us to situate class size within a constellation of factors that affect and are affected by classroom interactions. It is within these interactions that learning is fostered. Using a conception of quality that focused on instructional interactions provided a window to how class size reduction can make a difference in the daily lives of teachers and students. This approach rests on the foundation of a growing literature on classroom quality and interactions that describes learning contexts across the nation. It has proven to be a powerful metric for gauging the distinctive characteristics that promote achievement.

CSR and Quality are Complex Constructs

By examining class size reduction and classroom quality synergistically, we have explored the question: How do class size reduction and classroom quality interact to produce learning opportunities? This is important because the two co-occur in classroom contexts and can potentially work together. A first step was, to define operationally what we mean by "quality." In prior research, teacher quality has been discussed as a set of teacher qualifications, beliefs, dispositions, and effectiveness to raise test scores (Kennedy, 2008). This conceptualization is insufficient to understand how teachers interact with students. Moreover, because multiple teachers often interact with students, we must widen the lens to include classroom quality to gain a more inclusive understanding of children's day-to-day learning experience.

Using SAGE as the context for this paper adds a new layer to research on class size reduction because the program provides flexibility for schools to use reduced class size in various treatments. These include classes with fewer students, classes with more teachers, or classes where teachers are added to teach core subject areas. This variation in implementation highlights how imagining CSR as a reform that simply lowers class size is an oversimplification. CSR is not an uncomplicated input—within each of the treatments we observed, we found a range of quality ratings. CSR and quality are too complex to be considered in an input-output model.

The Importance of Using a Multi-Method, Context-Focused Approach

Disconnected knowledge of structural inputs, CLASS ratings, or standardized test scores cannot yield the type of nuanced, contextually-rich representations needed to understand classrooms. Taken together and paired with qualitative field notes and interviews, these data

reveal how classrooms could better utilize the resource provided by a CSR reform, regardless of configuration or treatment.

Variation in practice is not in and of itself an "ah-ha!" finding. In fact, it is absolutely congruous with the finding of variation in classes nationwide (NICHD, 2005). However, in this instance, variation illustrates the idea that class size alone does not facilitate an understanding of classroom quality. It lends credence to those who have asserted that class size is a contextual variable that must be considered in relation to other, equally important factors (Blatchford, 2003b; NICHD, 2005). By examining measures of quality in different forms of implementation of a large CSR program, we can begin to identify practices that engage the resource of class size reduction productively, as illustrated through the vignettes presented in this article.

Lessons Learned

Classrooms with high CLASS ratings provided emotional support, classroom organization, and instructional support to help students learn. This was accomplished organizationally in the 15:1 classrooms through practices that helped students develop strategies independently so that the teacher could provide targeted small group and individual instruction. This further reduced the instructional groupings so that children had more opportunities to interact with the teacher and with each other. It was also supported through instruction that was specific to the needs of particular children rather than to a generic curricular program.

Classroom organization in Ms. Helman's room exemplifies instruction that is tailored to the interests and needs of students, making it relevant and closely matched to their instructional level. The high quality feedback provided in each scenario was possible because the teacher listened carefully to students and supported them with relevant skill building. The contrast of quality in Mrs. Howard's classroom shows a lack of instructional differentiation, limited feedback, and management-related interruptions. This is an example of missed opportunities, with lessons that were not aligned to student needs, close-ended questions, and fractured interactions.

In the higher quality team-taught classrooms, the successful practices of the smaller classrooms (such as linking content to student experience and constant, iterative feedback loops) were accompanied by teacher familiarity with content and each other. Constructive teaming situations provided a seamless set of resources—even with more students. Mrs. Giles and Mrs. Martin were akin to ballroom dancers who knew the music and each other's strengths, bringing experience to bear on their interactions with students. Sometimes they danced together and sometimes apart, working to increase interactions with their students. In contrast, Farmington did not nurture the development of teaching partners. The constantly changing teams could not synchronize their practice to make the most of having two professionals in the classroom. Their instruction was more formulaic and less attuned to students' needs and students were less engaged.

Class size, CSR configuration, teachers' backgrounds, and administrative choices are potential enablers of promising instructional practices and classroom quality. Our vignettes are exemplars of the 27 classrooms we studied over 3 years. In each case, the typical variables used in input studies (class size or teacher characteristics) would have provided us with little insight into students' learning environments. We use these examples to suggest that what is done through class size reduction makes a difference for student learning. Smaller classes make

certain things possible—but those things are also linked to curricular mandates, educator beliefs, physical space, and the dynamics of classroom practice.

Strengths and Limitations

Our multi-method approach examined the synergy of class size reduction and classroom quality, providing greater insight than separately examining them would have yielded. The research we have presented here built on 3 years of fieldwork in 9 schools. These schools were chosen because they represented variation in attributes that we hypothesized were related to school success. The Classroom Assessment Scoring System is theoretically and empirically well-developed and proven in practice. Given the diversity of data collected, generated, and analyzed, we feel confident that our assertions are triangulated across sources, time, and focus.

The design of the study brings together measures of students aggregated at the school level, classroom quality, observations and interviews that link classroom and school context,. Some might see this approach as mixing levels of analysis and muddying the potential inferences to be made. We recognize this concern but argue that this is a matter of paradigmatic perspective and that no amount of controls would address this critique.

The structure of the SAGE program, with its multidimensional design, could be seen as complicating our analysis. The other pillars could be seen as alternative factors in the SAGE experience. The cost of staffing virtually always exceeded funds provided by the state; schools typically had to find additional funds to cover the cost. In addition, many informants told us that they had never heard of the other program elements (Graue, Rauscher, & Sherfinski, 2008). For these reasons, we feel that practice related to the other pillars is attributable to the existing variation among schools.

All research has limitations. Some key constraints in this work have come out of the need to design a study that used the existing conditions in schools. Not wanting to impose additional tests on schools, we did not have access to classroom-level measures of student achievement, nor is there an assessment instrument that was standardized across the six school districts we studied for the K-2 grades. Therefore, we could not measure the direct achievement effect of particular teachers or classroom quality. Third and fourth grade state achievement tests represent an aggregate of the K-3 schooling experience. In addition, we do not have comparison groups that could be seen as a control for treatments like class size or quality. This has been a particular issue in this evaluation as 1) SAGE is so widespread, and 2) it serves most of the schools educating students living in poverty in Wisconsin. Any comparison group would certainly have other attributes that could account for differences in practices or achievement.

Our sample, which is small by quantitative standards but large by qualitative standards, does not allow direct generalization to either the broader SAGE population or to the population of schools at large. Instead, the sample provides what Stake (1995) calls naturalistic generalizability, a process by which individuals can generalize from one experience to another, or transferability (Lincoln & Guba, 1985), which allows connections to broader literature showing considerable variation in quality in U.S. classrooms that is not completely interrupted through class size reform (NICHD, 2005).

Conclusion

Following Ball and Forzani (2007) we found that careful examination of the underlying processes in these classrooms provides a better sense of the mechanisms that produce the

patterns of quality. Class sizes afforded by participation in the SAGE program form a foundation for quality practice that must be leveraged through authentically nurtured emotional connections, thoughtfully planned and carefully carried out organizational strategies, and precise, targeted, instructional supports. The teachers in the high quality classrooms in our sample were able to take advantage of the opportunities of class size reduction in diverse ways, regardless of classroom configurations or treatments. What they had in common was careful planning calibrated to the needs of specific children—needs that encompassed both the social and academic dimensions of learning.

The considerable variation in classroom quality in our sample mirrors variation found in the broader education community (NICHD, 2004). The substantial investments in CSR (which in the case of the 2006-2007 SAGE implementation was $98,588,000) are insufficient alone to broadly enact defined conceptions of quality. However, this does not disprove the value of class size reform. Instead it shows that understanding strategies for improving education opportunities is more complex than investing in simple inputs such as class size. Curiously, the structures of the SAGE program, as initially designed, anticipated this complexity and included attention to curriculum, professional development, and the strengthening of home-school connections by making the school building a hub for social services and resources. Because the funding of SAGE typically does not cover even the staffing component (Graue et al., 2008), this considerable injection of funds needs proximal support to leverage the intended changes. The potential of class size reduction could be realized through a multifaceted approach that brings together knowledge of the specific contexts of implementation and practices that promote classroom quality. By bringing these two sources of knowledge together we can provide important guidance to educators who have been given an instructional resource that is highly valued but poorly understood.

Notes

Support for this work came from a grant from the Wisconsin Department of Public Instruction (DPI). The ideas presented represent the positions of the authors, not the DPI. We are grateful to the educators who contributed their time, expertise, and guidance to this project. In addition, the work benefitted from the feedback provided by ESJ editor Betsy Davis as well as Peter Blatchford and anonymous reviewers.

Correspondence concerning this article should be addressed to Beth Graue, Wisconsin Center for Education Research, 1025 W. Johnson, Madison, WI 53706. E-mail: graue@education.wisc.edu.

1. A professional videographer accompanied a researcher to the classroom and taped classroom interactions with two cameras while the teacher wore a wireless microphone to record conversations. These videos were then transferred to a digital format and viewed multiple times and coded.

2. To be certified, raters attend a two-day training that introduces the instrument and its research support. Participants view videos that illustrate each dimension, practice coding five videos as a group, and perform reliability testing on an additional five videos. To be certified, each coder must attain 80% reliability with master coders on reliability training videos and correctly rate 2 of 5 examples on each dimension.

3. In one classroom, academic practice was available across only 3 observation sessions. For several classrooms, there were 5-6 observation cycles and the 4 most representative were included in the scoring.

4. The actual CLASS data sheets are standardized with space for notes and scoring for each of the 10 CLASS dimensions. This vignette analysis of Ms. Helman's Morning Meeting shows a CLASS rater's evaluation of practice for a portion of one 30-minute cycle. In CLASS, the ratings are assigned from the presence of behaviors distributed over the full observation cycle; the scores listed for the vignette represent a small section of a cycle. As CLASS involves analyzing multimodal cues, it is done from real classroom observation or videotape. Field note analysis is undertaken here to demonstrate the CLASS constructs for our readers.

5. School performance on state standardized tests was the only measure available across all six study school districts. Grade four was chosen as a post SAGE measure of student achievement given its administration in fall of grade 4 (SAGE served students K-3).

6. This relationship is not causal; it is presented as a validity marker.

7. In whip share participants consecutively answer the question quickly.

References

Ball, D. L., & Forzani, F. (2007). What makes education research "educational?" Educational Researcher, 39(9), 539-540.

Biddle, D.L., & Berliner, D.C. (2002). Small class size and its effects. Educational Leadership, 59(5), 12-23.

Blatchford, P. (2003a). A systematic observational study of teachers' and pupils' behaviour in large and small classes. Learning and Instruction, 13(6), 569-595.

Blatchford, P. (2003b). The class size debate: Is smaller better? Maidenhead: Open University Press.

Blatchford, P., Bassett, P., & Brown., P. (2005). Teachers' and pupils' behavior in large and small classes: A systematic observation study of pupils aged 10 and 11 years. Journal of Educational Psychology, 97(3), 454-467.

Burke, D.L. (1996). Multi-year teacher/student relationships are a long-overdue arrangement. Phi Delta Kappan, 77(5), 360-361.

Cassidy, D. J., Hestenes, L. L., Hansen, J. K., Hedge, A., Shim, J., & Hestenes, S. (2005). Revisiting the two faces of child care quality: Structure and process. Early Education & Development, 16(4), 505-520.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. Psychological Science and the Public Interest, 2(1), 1-30.

Englehart, J. M. (2006). Teacher perceptions of student behavior as a function of class size. Social Psychology of Education, 9(3), 245-272.

Englehart, J. M. (2007). Discourse in a small class. Education, 35(1), 83-97.

Erickson, F. (1990). Quantitative Methods Qualitative Methods: A Project of the American Educational Research Association (Research in Teaching and Learning, Vol. 2) by Robert Linn and Frederick Erickson. New York: Macmillan.

Finn, J. D. & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. American Educational Research Journal, 27(3), 557-577.

Finn, J. D., Pannozzo, G. M., & Achilles, C. M. (2003). The "why's" of class size: Student behavior in small classes. Review of Educational Research, 73(3), 321-368.

Graue, E., & Rauscher, E. (2009). Researcher perspectives on class size reduction. Education Policy Analysis Archives, 17(9).

Graue, E. & Rauscher, E. (2009). Reclaiming assessment through accountability that is "just right." Manuscript submitted for publication.

Graue, E., Rauscher, E., & Sherfinski, M. (2008). SAGE implementation & classroom quality. Madison, WI: Wisconsin Center for Education Research.

Graue, M. E., Hatch, K., Rao, K., & Oen, D. (2007). The wisdom of class size reduction. American Educational Research Journal, 44(3), 670-700.

Graue, M. E., & Walsh, D.J. (1998). Studying children in context: Theory, methods, and ethics of studying children. Thousand Oaks: Sage Publications.

Grissmer, D. (1999). Class size effects: Assessing the evidence, its policy implications, and future research agenda. Educational Evaluation and Policy Analysis, 21(2), 231-248.

Hamre, B., & Pianta, R. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? Child Development, 76(5), 949-967.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 U.S. early childhood and elementary classrooms. Foundation for Childhood Development. Report retrieved on January 30, 2008, from http://www.fed-us.org/resources/resourcesshow.htm?doc id=507559.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn: Children's pre-academic achievement in pre-kindergarten programs. Early Childhood Research Quarterly, 23(1), 27-50.

Kennedy, M. (2008). Sorting out teacher quality. Phi Delta Kappan, 90(1), 59-63.
Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from project STAR. Elementary School Journal, 108(4), 275-292.

LaParo, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. Elementary School Journal, 104(5), 409-426.

Lincoln, Y.S. & Guba, E. (1985). Naturalistic inquiry. Thousand Oaks, CA: Sage Publications.

Maxwell, J.A.(1996). Qualitative research design—an integrative approach. Thousand Oaks, CA: Sage Publications.

Molnar, A. & Zmrazek, J. (1994). Improving the achievement of Wisconsin's students: Urban Initiative Task Force recommendations and action plan. (Bulletin No. 95079). Madison, WI: Wisconsin Department of Public Instruction. Office of Policy & Budget.

National Institute for Child Health & Development. (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? Developmental Psychology, 40(5), 651.

National Institute for Child Health & Development. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. Elementary School Journal, 105(3), 305-323.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2004). Do minorities experience larger lasting benefits from small classes? Journal of Educational Research, 98(2), 94-100.

Pate-Bain, H., Achilles, C. M., McKenna, B., & Zaharias, J. (1992). Class size makes a difference, Phi Delta Kappan, 74(3), 253-256.

Pianta, R.C., Belsky, J., Houts, R., Morrison, F., & The National Institute for Child Health and Human Development (NICHD) (2007). Opportunities to learn in America's elementary school classrooms. *Science, 315*, p. 1795-1796.

Pianta, R. C., LaParo, K., & Hamre, B. (2008). CLASS Classroom Assessment Scoring System manual K-3. Baltimore: Paul H. Brookes.

Pianta, R. C., LaParo, K., 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.

Rimm-Kaufman, S. E. (2006). Social and academic learning study on the contribution of the responsive classroom approach. Northeast Foundation for Children, Inc. Report retrieved on July 2, 2009 at http://www.responsiveclassroom.org/pdf_files/sals_booklet_rc.pdf

Rimm-Kaufman, S. E., LaParo, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children. Elementary School Journal, 105(4), 377-394.

Schensul, S. J., Schensul, J. J., & LeCompte, M. D. (1999). Essential ethnographic methods. Observation, interviews & questionnaires. Walnut Creek, CA: Alta Mira

Smith, P., Molnar, A., & Zahorik, J. (2003). Class-size reduction: A fresh look at the data. Educational Leadership, 61(1), 72.

Stake, R. (1995). The art of case study research. Thousand Oaks, CA: SAGE.

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

Webb. M. L., & Meyer, R.H. with Gamoran, A. & Fu, J. (2004). Participation in the Student Achievement Guarantee in Education (SAGE) program & performance on state assessments at grade 3 & grade 4 for three cohorts of students—grade 1 students in 1996-97, 1997-98, & 1998-99. Madison, WI: Wisconsin Center for Education Research.

Wilson, H. K., Pianta, R. C., & Stuhlman, M. (2007). Typical classroom experiences in first grade: The role of classroom climate and functional risk in the development of social competencies. Elementary School Journal, 108(2), 81-96.

Zahorik, J., Halbach, A., Ehrle, K., & Molnar, A. (2003). Teaching practices for smaller classes. Educational Leadership, 61(1)