Early Childhood Research Quarterly 24 (2009) 198–211

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Keeping their attention: Classroom practices associated with behavioral engagement in first grade mathematics classes in China and the United States[☆]

Xuezhao Lan[a,*], Claire Cameron Ponitz[a,1], Kevin F. Miller[a], Su Li[b], Kai Cortina[a], Michelle Perry[c], Ge Fang[b]

[a] Combined Program in Education and Psychology, University of Michigan, 1044 D East Hall, 530 Church Street, Ann Arbor, MI 48109, USA
[b] Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[c] University of Illinois Urbana-Champaign, IL, USA

ARTICLE INFO

Article history:
Received 5 January 2008
Received in revised form 3 March 2009
Accepted 4 March 2009

Keywords:
Behavioral engagement
Cross-cultural comparison
First grade
Mathematics instruction
Chinese children

ABSTRACT

This study investigated variation in students' behavioral engagement across mathematics classes in China and the United States. Student behavioral engagement was examined along with two aspects of the classroom (group size and teacher instructions given about classroom behavior). Video observational data were collected and coded over 1051 time intervals in 35-minute mathematics sessions in Chinese classrooms ($n = 8$) and comparable American classrooms ($n = 7$). Latent growth analyses revealed that overall, behavioral engagement declined over time, although the drop-off was dramatically sharper in American classrooms relative to Chinese classrooms. In addition, larger group size and the timing of teacher instructions (given before versus after the behavior) were significantly associated with increased engagement. This study revealed compelling cultural differences as well as patterns in student and teacher behaviors associated with students engaging in on-task behaviors in the classroom. Implications for ways to promote effective classroom behavior are discussed.

The mathematics achievement of American students has received increasing attention as educators and policymakers strive to help the United States remain competitive in an emerging global economy (Heckman, Stixrud, & Urzua, 2006). International comparisons typically have found that students in Asian countries show the highest levels of average mathematics performance (Akiba, LeTendre, & Scribner, 2007; Beaton et al., 1996; Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985; Robitaille & Garden, 1989; Stevenson & Stigler, 1992). Achievement differences reported in international and comparative studies have motivated policy initiatives focused on mathematics education reforms in the United States (Romberg, 1997, 1999).

Differences in achievement almost certainly reflect differences in teaching and learning processes. Practices exist within a larger context of beliefs and resources relevant to education; nonetheless, investigating educational practices in nations with strong mathematics performance can be a useful tool for understanding factors that promote achievement in this critical subject area (Stigler, Gallimore, & Hiebert, 2000). We will review some of the known differences in early teaching

☆ This article is based upon work supported by National Science Foundation Grant 0089293. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

* Corresponding author. Tel.: +1 734 330 0180; fax: +1 734 615 2164.
E-mail address: xuezhao@umich.edu (X. Lan).

1 Claire Cameron Ponitz is now at the University of Virginia, Center for Advanced Study for Teaching and Learning.

0885-2006/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2009.03.002

and learning practices between China and the United States before describing a study that looked at classroom practices in China and the United States during a mathematics lesson related to a critical behavioral predictor of achievement – students' engagement during a lesson. The present study focuses on young children (first grade elementary school children) who have begun systematically learning mathematics and developing learning and regulating skills in China and the U.S.

1. Cross-cultural differences in mathematics education

Substantial differences in achievement in East Asian countries and the United States have been systematically linked to differences in how families, teachers, and society-at-large treat the teaching and learning of mathematics. For example, American parents and teachers are more likely to attribute mathematics competence to ability, whereas Japanese and Chinese adults tend to believe that learning comes with effort and persistence (Stevenson, Lee, Chen, & Stilger, 1990). Since 1995, the Third International Mathematics and Science Study (TIMSS), which compares competence in different societies, has documented cultural differences in attitudes, practices, and achievement in mathematics. In general, East Asian countries emphasize explicit mathematics instruction, promote a deeper conceptual grounding than American students typically receive, and encourage students to put significant and sustained effort into learning math (Ma, 1999; Miller, Kelly, & Zhou, 2005; Stevenson & Stigler, 1992). These differences are connected to multiple factors, both distal and proximal to the classroom. East Asian societies are known to be collectivistic and Western societies individualistic (e.g. Nisbett, 2003), such differences may influence both societies at curriculum and classroom levels. Asian countries have a more centralized curriculum system (Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999), and those curricula are more focused, integrated, and coherent across schools and grade levels. In contrast, the U.S. system for adopting curricula is heavily subject to marketing pressures, with decisions made primarily at the district and school level. This contributes to American mathematics curricula being less authoritative and less consistent across years and among districts, compared to the single, national curriculum taught in Asian countries (Cohen & Spillane, 1992; Peak et al., 1996).

Difference among American curricula may underlie the widely observed variability in teaching in the United States (National Institute of Child Health and Human Development Early Child Care and the Research Network [NICHD ECCRN], 2002). For example, one study found substantial variation in amount of teacher organization as well as transition, teacher-, and child-directed instruction (Cameron, Connor, & Morrison, 2005). Classroom-based work has also revealed American and Asian cross-cultural variation in instructional practices, including how teachers promote conceptual mathematics understanding (Hiebert & Stigler, 2000; Perry, 2000; Stigler & Perry, 1988; Stevenson & Lee, 1995; Stigler & Hiebert, 1999). For example, Yang and Cobb (1995) found that Chinese teachers encouraged students to construct composite, multiunit numerical conceptions (e.g., 14 is composed of a unit of ten and four ones) and to justify their solutions (The Chinese language also facilitates such representation as 14 is pronounced as "ten-four"). On the contrary, U.S. teachers in the study encouraged students to construct unitary concepts (e.g. 14 is the number after 13) with little justification. Furthermore, compared to American teachers, Chinese teachers engaged in more extended discourse, such as using a student's answer to a question to begin a larger discussion about the mathematical algorithms, rules, and reasoning needed to find that answer (Schleppenbach, Perry, Miller, Sims, & Fang, 2007). Finally, East Asian students spend substantially more time than American students studying mathematics (Stevenson, Lee, Chen, & Lummis, 1990).

Taken together, these findings reveal significant differences in how mathematics is taught and learned in the United States and Asian countries, including China. But effective learning involves factors that extend beyond the curriculum or formal instructional strategies. Another key is student engagement, the importance of which was highlighted in a report on mathematics learning from the National Research Council (Kilpatrick, Swafford, & Findell, 2001). Student engagement, as well as the strategies that teachers use to ensure that students remain engaged, can also be a central predictor in accounting for student achievement (Fredricks, Blumenfeld, & Paris, 2004; Greenwood, Horton, & Utley, 2002).

2. Teacher organizational instructions

The National Council of Teachers of Mathematics (1991, 2000) has called for teachers to use organizational strategies, including instructions that allow students to effectively learn mathematics. A significant body of research (mostly in the United States) has documented how teacher organization for instruction sets the stage for effective classroom functioning by helping students become behaviorally engaged and regulate their own actions (Bohn, Roehrig, & Pressley, 2004; Brophy, 1985; Brophy & Good, 1986; Cameron et al., 2005; Evertson, Emmer, Sanford, & Clements, 1983; Pressley et al., 2001; Sanford & Evertson, 1983). Organizational strategies include teacher efforts to preview classroom activities, present instructions about their completion, and provide clear expectations for student behavior (Anderson, Evertson, & Emmer, 1980). Teacher practices such as giving clear instructions about tasks in advance have been associated with greater behavioral engagement (Carta, 1991). Such efforts are thought to help prevent behavioral difficulties and curtail potential disciplinary distractions by creating a predictable, organized learning environment (Brophy, 1988).

Though limited, the existing cross-cultural comparison literature indicates that Chinese teachers, when compared with American teachers, are able to use their teaching time more effectively for student learning and to develop better-organized whole-class instruction– even in their typically large classes (Yang & Cobb, 1995). A key indicator of organizational strategies on the part of the teacher may be the timing of instructions. Instructions given before students begin a task can provide them with a model of what they should be doing, and in turn promote self-regulation. Instructions given after a task begins may reflect either shortcomings in the initial instructions or the need to get children back on task. We therefore looked at the kinds of instructions teachers used to encourage or request behavioral engagement, including whether instructions were provided *before* or *after* a task within two countries (China and the United States). We predicted that instructions given before a task would predict increased behavioral engagement, and expected that Chinese teachers might give a greater proportion of instructions before a task commences, compared with their U.S. peers.

3. Student behavioral engagement within and across cultures

In classroom contexts, behavioral engagement signals children's overt involvement in learning tasks, or the extent to which behavior aligns with teacher expectations (Fredricks et al., 2004; Zanolli, Daggett, & Pestine, 1995). On-task behavioral engagement is observable in the classroom, especially with younger children, and includes active behaviors, such as asking or answering questions; and passive behaviors, such as listening and writing. Research indicates that behavioral engagement underlies students' adaptability to the demands of the classroom setting (Ladd, Birch, & Buhs, 1999). As such, engagement is a vital contributor to academic achievement and has become an outcome of interest in its own right (Fredricks et al., 2004; Greenwood et al., 2002). Foundational early work has revealed that engaged time in school is positively associated with student achievement, with high-achieving students spending up to 45% more time engaged in academic activities than low-achieving students (Evertson, Anderson, Anderson, & Brophy, 1980; Stallings, 1975). Since then, both observational and experimental work has connected greater engagement and related behavioral indicators with a host of positive outcomes, including social functioning and mathematics (Blair & Razza, 2007; Greenwood, 1991; McClelland, Acock, & Morrison, 2006; Peterson & Fennema, 1985; Ponitz, Rimm-Kaufman, Curby, & Grimm, in press).

Cultural differences in classroom behavior related to engagement have also been documented, and these co-occur with differences in the opportunities teachers provide. One inquiry found that Chinese teachers were more likely to vary their instructional tasks to hold student attention and encourage students to respond in a rapid manner compared to their U.S. peers (Stevenson & Lee, 1995). In another study (Clarke-Stewart, Lee, Allhusen, Kim, & McDowell, 2006), Korean children exhibited greater self-reliance and sustained attention compared to their American counterparts. However, children in American classrooms interacted more with their peers and had access to more varied materials for gross motor activities and socio-dramatic play than did Korean children.

4. Engagement and classroom settings

Different opportunities for engagement may be associated with the setting in which instruction occurs, although we are not aware that this has been examined cross-culturally. Examples of different settings include whole-group work, small-group work, or individual work. Whole-group settings may allow the teacher to command the attention of all children, who direct their attention and activity based on teacher instructions rather than on their own plans. Gump (1969) found greater student involvement during large-group recitation activities and when the teacher was presenting work, as compared with independent seatwork. In general, American children show higher levels of behavioral engagement when the teacher is present, versus when they are left to work on their own (Rimm-Kaufman, La Paro, Downer, & Pianta, 2005).

Cultural differences in classroom settings have been reported in how teachers deliver instruction, attempt to engage students, and provide feedback. Stigler and Perry (1988) found that Chinese teachers were more likely to use whole-class settings for instruction, whereas U.S. teachers were more likely to use small-group or individual instruction. Higher average overall class sizes in China compared to the United States may help contribute to this difference in practice; small-group settings may be easier to manage with fewer students in one class. In a study of the early years of school in England, small classes were shown to provide more opportunity for individualized feedback (Blatchford, Moriarty, Edmonds, & Martin, 2002). Even so, American students spend the majority of their time in whole-group settings (Pianta, La Paro, Payne, Cox, & Bradley, 2002). However, even in whole-class settings, Chinese teachers tend to provide individualized feedback and have rich conversational exchanges, which may indicate a difference among Asian and European/American settings (Schleppenbach et al., 2007).

5. Why study China?

The current study examined key elements of mathematics classrooms in China and the United Sates, including teacher practices beyond mathematics content instruction, such as organizational instructions for student behavior, and students' behavioral engagement in mathematics classes. We fully acknowledge that teaching is culturally embedded, and admit that our investigative strategy will not allow us to differentiate among multiple variables, which may have cultural or possibly, political, roots (Santagata & Stigler, 2000). Comparing different countries can nonetheless help us illuminate potentially useful practices in both countries. We chose to focus our cross-cultural lens on classroom practices in China for three main reasons:

First, as we have discussed, Chinese students outperform their U.S. counterparts. This advantage has been documented in multiple skill areas such as counting and understanding place value (Miller & Stigler, 1987; Miura, Chungsoon, Chang, & Okamoto, 1988), as well as calculation and mental mathematics (Brenner, Herman, Ho, & Zimmer, 1999; Geary, Bow-Thomas, Fan, & Siegler, 1993).

Second, China's achievement in mathematics is all the more striking given its relative disadvantage with regard to financial resources, compared to fully industrialized nations. Chinese educators typically have less funding than the United States, and scholars have argued these funds have been less equitably divided than in the U.S. (Hannum & Wang, 2006; Stevenson & Stigler, 1992). In addition, China and the United States share challenges not faced by small countries, due to their large populations and diverse geographical regions (Chen & Stevenson, 1995; Ma, 1999; Miura et al., 1988; Stevenson, Lee, Chen, & Stilger, 1990; Stevenson, Lee, Chen, & Lummis, 1990). However, despite these factors, Chinese students demonstrate strong academic achievement, often achieving at higher levels compared to American students with access to more educational resources.

Third, classroom practices, including those during mathematics instruction, differ in systematic ways. Compared to their American counterparts, Chinese teachers have been shown to make more efficient use of their class time, and engage students in inquiry by using whole-class pedagogical techniques (Linn, Lewis, Tsuchida, & Songer, 2000; Perry, 2000; Stevenson & Lee, 1995; Stigler & Hiebert, 1999). However, few studies have directly examined behavioral engagement during mathematics instruction. Using a cross-cultural lens to examine teaching practices associated with high levels of engagement might provide insights into factors associated with effective mathematics learning.

6. Research questions and hypotheses

First, what is the nature and extent of cultural differences in teacher instructions about behavior, and behavioral engagement in large-group or small-group settings? American classrooms were expected to show a greater level of variability compared to the Chinese classrooms on all the variables. Chinese teachers were expected to give more proactive behavioral instructions (instructions given prior to student behaviors as opposed to reactive instructions given after student behaviors), by giving more complete instructions before children started a task. Chinese teachers were also expected to utilize a greater proportion of large-group activity settings. It was also hypothesized that Chinese students would demonstrate higher levels of behavioral engagement compared with American students.

Second, what are predictors of students' on-task behavior? Specifically, being in a Chinese classroom, participating in teacher-directed large-group settings, and receiving proactive versus reactive instructions were expected to be associated with greater behavioral engagement.

7. Method

7.1. Participants

7.1.1. Schools

Four schools (one private school and three public schools) in central Illinois in the Midwest United States and three public schools in Beijing, China (a large urban area) were recruited to participate in this study. In the American sample, three schools were from a university town area (including two public schools and the private school); the other one was located outside of the university town in a rural area. Three classes were recruited from the rural school. In the Chinese sample, all schools were public schools from three different districts in Beijing. These samples were recruited by contacting schools in both areas that were deemed by local experts to be typical for the area (i.e., not exceptional, or key schools, etc.). With the permission of the school principal, all the first grade teachers in those schools were contacted. If teachers agreed to participate, they were videotaped when they taught a lesson on place value (a topic taught in both countries at approximately the same time in the school year).

7.1.2. Teachers and classrooms

Fifteen first-grade teachers (7 in the United States, 8 in China) consented to participate in this study.

All the U.S. teachers were female and Caucasian. Five of these teachers had experience teaching in middle-sized Midwestern U.S. cities while others had experience teaching in rural settings. They did not report to specialize in teaching a certain subject. All but one teacher worked in public schools. Lessons were videotaped in the spring semester of the students' first grade year. The average observed class size on the day of videotaping was 19 (*SD* = 1.72). The average length of lessons was 35 min (*SD* = 13.03). In three out of seven of the classrooms, each student was assigned to his or her own desk; in the rest of the classrooms, small groups of students (4 to 8) were assigned to share several large tables.

The Chinese teachers were all female and of Chinese origin. All these teachers specialized in elementary mathematics and taught multiple grade levels in public schools in Beijing. Lessons were videotaped in the spring semester of the students' first-grade year. The average observed class size on the day of videotaping for the Chinese class was 40 students (*SD* = 1.99). The average length of the Chinese classes was 39 minutes (*SD* = 2.43). In five classrooms, each student was assigned to a single desk; in the remaining three classrooms, small groups of students (5 to 6) were assigned to share several large tables.

7.1.3. Students

In China, students were of Chinese origin (100%). Students from the United States were 90% Caucasian, American-born children; the remaining 10% were of African American ethnic backgrounds and some of Asian backgrounds. Students from both countries represented the background of the respective local population. It is possible that some Asian American families parented in ways consistent with the Asian rather than American culture, though we did not measure children's home environments. However, the percentage of Asian American children in our U.S. sample was small so we did not consider student ethnic background in our analyses. The average age of first graders was 6.5 years in the U.S. and 6 years in China.

The majority of students' parents consented for their children to participate in another study collecting child data, with a few exceptions. We avoided videotaping children who did not consent to participate. All the video coding and analyses for the present study were done at the level of the whole class.

Although the schools in each nation were drawn from distinctly different areas, available evidence on student achievement and reputation indicates that participating schools in the two samples could be considered average to above average achieving schools. In addition, although we did not measure family income directly, we made an effort to recruit families from comparable schools. Students from schools in both sites came from a range of family backgrounds. In the small university town, students were from a mixture of university families and rural families. In the urban areas in Beijing, most students were from either middle-class or working class families.

It should be clear that neither sample could represent the diversity of two large and complex societies. Our goal is to identify potentially valuable educational practices and to understand the dynamics of student behavioral engagement across very different settings, rather than to evaluate the educational systems of China and the U.S. Finally, the students and teachers in our study were not aware of our hypotheses. They were told that goals of the study were to understand how students learn mathematics, and to examine how videotaped data can be used to train teachers.

7.2. Collecting observational data in classrooms

Observing the daily activities of classrooms has become an important method of studying teaching and learning. Observations have been increasingly used to supplement or replace teacher survey measures, which may not reflect actual classroom practices (Blatchford et al., 2002; Pianta et al., 2005). Studies have shown that sampling time intervals in classrooms yields relatively robust data representative of typical classroom happenings (Rimm-Kaufman et al., 2005). In the present study, math sessions in 15 first-grade classrooms were videotaped and coded for teacher instructions about behavior, classroom organization, and student behavioral engagement (classroom-level). Data were based on a single, one-day observation. By time-sampling the dependent and independent variables every 30 seconds, data provide the temporal dynamics of student engagement as well as associations with teacher behavior and classroom setting.

7.2.1. Coding procedure

Mathematics lessons in each classroom were videotaped with two cameras; one directed at the teacher, and the other at the students. Videotapes were coded at 30-second intervals, resulting in 1051 total time intervals (in a few instances, we did not stop the video at the exact 35th minute to allow teachers to finish their sentences; in a few other cases, the length of the classes was less than 35 minutes). Because the length of each class varied somewhat, the first 35 minutes of each session were coded, which captured most of the math lesson in both countries, a strategy employed in other observational research (Rimm-Kaufman et al., 2005). Thus there were 70 intervals per classroom, coded with a time-sampled strategy. These codes capture pre-determined behaviors and activities occurring in each 30-second interval, providing information on the nature and variation in teacher instructions, classroom setting, and task-related behaviors over time.

7.2.2. Coding system development

The coding scheme was developed by a group of American and Chinese researchers, including an American-born university professor, a Chinese professor, and a Chinese doctoral student who has lived in both countries. Goals were to identify behaviors occurring in both countries that might be associated with student engagement. Since we were particularly interested in behaviors relating to student engagement and teacher organization, we built upon previous literature in this area. We identified and operationalized related aspects of classroom organization (e.g. group size) and teacher regulating instructions. To select instructions, all regulatory instructions from ten comparable videotapes in both countries were documented, categorized, counted, and ranked according to their frequency by two educational psychologists from the University of Michigan and the Chinese Academy of Sciences. The five most frequent instructions in each country were chosen for final analysis, including eight types of oral instructions and two types of gestural instructions.

7.2.3. Training

Two English–Chinese bilingual coders were trained in one full-day coding workshop. Both coders were graduate students in psychology and were blind to study hypotheses. Both coders were of Chinese origin and had significant experience living in the United States. Each coder coded about half of the Chinese classroom observations and half the American observations. Both coders were trained and tested on coding, supervised by researchers from both institutions.

Prior to the training workshop, coders carefully reviewed code descriptions and viewed videotapes of classrooms. During training, they were instructed with two detailed example videos for coding and questions that might arise during the coding process. After the workshop, coders conducted pilot observations and coded one to two additional videotaped cases. Finally, coders were required to pass a videotaped reliability test involving two 44-minute cycles for behavioral coding. Coders obtained an 80% match (Cohen's Kappa) with each other and maintained this level of reliability on a reliability retest with two classrooms chosen randomly. Both coders passed at these levels on a reliability test before being certified to code the videos used for data analysis.

7.3. Codes for classroom activities and student behavioral engagement

For the time-sampled codes, each 30-second interval ($N = 1051$) was coded for three categories: *Classroom-level behavioral engagement, activity setting*, and *teacher regulatory instructions.*

7.3.1. Classroom-level behavioral engagement

Engagement was coded as a dummy variable at the classroom level, as either on-task or off-task. The goal was to identify intervals in which most of the students were engaged, and to differentiate from those in which a substantial number of students were off-task. Thus, "on-task" was defined when greater than the majority of students actively (e.g., discussing, asking or answering questions, a teacher asked students a question and most of the students answered/intend to answer the question) or passively (e.g., listening, reading, writing) engaged in a classroom task. To be considered on-task, 70% of students[2] had to be observed engaged for longer than 20 seconds in the 30-second interval. Otherwise, the interval was coded off-task (see also Rimm-Kaufman et al., 2005). Examples of off-task behaviors include engaging in laughing or free chat, sleeping, or looking around. In small groups, including in child-directed activities (for example, discussion among children or solving math problems all together), task-related conversations with classmates were considered on-task. In contrast, behaviors such as fighting with classmates, running around the classroom, or engaging in laughing or free chat were coded off-task.

7.3.2. Activity setting

Setting was coded as large group (with more than 6 students per group) or small group (with less than or equal to 6 students per group).

7.3.3. Teacher regulatory instructions

Teacher regulatory instructions eliciting on-task student behavior were defined as instructions that did not contain domain knowledge, such as drawing students' attention to a task. Both oral and gestural instructions were coded.

Oral teacher instructions included (1) "Stop!"; (2) "Pay attention"; (3) Social comparison such as "Who can finish first?"; (4) Classroom norms, for example, "Sit/raise your hand properly" (for students who sit inappropriately such as lying on the table, or do not raise their hands correctly); (5) "Use complete sentences when asking a question," (6) "Speak up," (7) "Raise your hand," and (8) "Sit down." The two types of gestural instructions were (9) "raise your hand" (i.e., the teacher raised her hand with her mouth closed, indicating that students should do the same before talking), and (10) "sit down" (i.e., the teacher made a "down" motion with her forearm).

In addition, each type of instruction was also coded as either proactive or reactive, defined as occurring before or after the student behavior (or misbehavior)[3]. Instances of instruction types were counted in each 30-second interval. The majority of the 30-second time slots contained only one instance of each type of instruction, although in several intervals, there was more than one instance of an instruction. Results were similar when we analyzed number of instances versus percentage of time slots.

8. Results

Our analytic plan was twofold. First, we examined the nature of coded classroom variables by comparing the means and variances of these variables across countries. The variables included student behavioral engagement, activity setting, and teacher instructions. Descriptive analysis revealed a striking difference between the behavioral engagement levels (probability of being engaged from the beginning to the end of a lesson) of our two samples (Fig. 1). Taken as a whole, the American classrooms show a relatively high level of engagement at the beginning of class, with a generally consistent drop over time. The Chinese classes start with a slightly higher level of engagement than the U.S. classes, and show only a slight decrease in engagement over time. In order to understand what factors might account for these differences, we then used latent growth modeling to examine factors that predict changes in the dynamics of student behavioral engagement across the lessons. These will be discussed in turn.

8.1. Cultural differences in behavioral engagement, activity setting, and teacher instructions

Table 1 shows the percentage of the 35-minute coded blocks in which each coded behavior occurred in each country, including behavioral engagement (on-task versus off-task behavior), activity setting by small or large group size, and teacher

[2] 70% was calculated based on the observable children. Although we had two camcorders in each classroom and did our best to follow all children in the classroom, sometimes we were unable to capture 100% of the children, due to movement.

[3] This is both at the individual and whole classroom level. For example, if a teacher instructed the whole class to "raise your hand" before answering a question, and then asked a student to answer a question, it was considered proactive instruction.

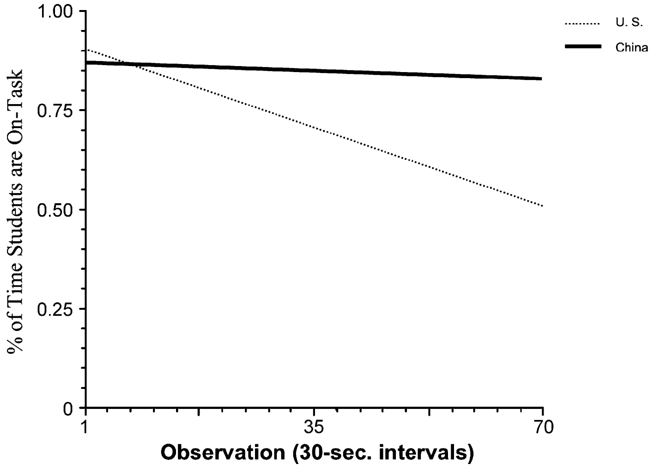

Fig. 1. Average % of time student on-task from the beginning to the end of a class in China and the U.S.

instruction by oral and gesture type. Greater variance on all the variables was observed in the American classrooms as compared to the Chinese classrooms (see Table 1). For example, variability in overall engagement in U.S. classrooms was significantly greater than in the Chinese classrooms ($SD_{U.S}$ = .37, SD_{China} = .05, Levene's $F(1,14) = 14.69$, $p < .01$).

8.2. Behavioral engagement

An analysis of variance showed that behavioral engagement was coded in a significantly higher percentage of time slots in Chinese classrooms (96%) compared to American classrooms (61%), $F(1, 14) = 7.54$, $p < .05$ (see Table 1).

8.3. Classroom setting

Teacher-directed large-group activities occurred during 58% of time slots in U.S. classes, compared to 93% of time slots in Chinese classrooms, and this difference was significant, $F(1,14) = 8.69$, $p < .01$.

8.4. Teacher regulatory instructions

To investigate the 10 different teacher regulatory instructions that were coded, these instances were summed across all the intervals. Then, the number of instructions by type (proactive or reactive) was summed. Teachers in the two countries differed in the amount and type of instructions they provided. Independent sample *t*-tests showed that Chinese teachers gave significantly more proactive instructions than American teachers, $t(1) = -5.48$, $p < .01$. Chinese teachers gave proactive instructions during 33% of time intervals, compared with 11% for U.S. teachers. In addition, the majority of the oral behavioral instructions, or 69% ($n = 269$, out of total instructions) given by Chinese teachers were proactive. This contrasted with 102 total instances of oral behavioral instructions in the United States, of which 48% were proactive.

Table 1
Average percentage of observations in each category.

	U.S.		China	
	Ave. %/class	SD	Ave. %/class	SD
Group size				
Large group	58	.32	93	.10
Small group	42	.32	7.0	.10
Student engagement				
Engaged	61	.37	96	.05
Disengaged	39	.37	4.0	.05
Instructions (out of total observations)				
Proactive	11	.06	33	.10
Reactive	12	.13	15	.04
Instructions (out of total instructions)				
Proactive	48	.22	69	.09
Reactive	51	.22	31	.09

Table 2
Average percentage of instructions out of total instructions.

	U.S.		China	
	Ave. %/class	SD	Ave. %/class	SD
Stop	39	0.23	7.7	0.06
Pay attention	21	0.19	27	0.09
Social comparison	1.5	0.12	20	0.01
Classroom norms	19	0.26	10	0.07
Use complete sentences when asking a question	1.5	0.04	4.4	0.02
Speak up	3.7	0.04	15	0.08
Raise your hand	9.0	0.35	7.1	0.04
Sit down	5.3	0.04	8.8	0.05

American teachers gave reactive instructions in 11% of intervals, which did not differ significantly from Chinese teachers (15%); however, U.S. teachers gave significantly more reactive instructions (52% out of total instructions) compared to proactive instructions (48%), $t(1) = 5.46$, $p < .01$. Gestural instructions followed the same trend, with larger differences across countries. Nearly 96% of the gestural instructions given by Chinese teachers were proactive, compared to 28% given by U.S. teachers.

The most frequent command in the U.S. was for students to stop doing something ("Stop", 39%, see Table 2 for a complete list of instructions), and 80% of this type of instruction occurred after the student behavior. Specific instructions were distributed relatively equally by type in China: "Pay attention" was the most frequent instruction, which accounted for 27% of all behavioral instructions; Social comparison accounted for 20% of all behavioral instructions in China (versus 1.5% in the U.S.). Almost all the above instructions were (99%) proactive instructions in China, occurring before the student behavior.

Thus, there were striking differences in overall levels of student behavioral engagement in our two samples, with Chinese students showing higher and more consistent levels of engagement than their U.S. peers. Teacher instructional practices also differed, with the Chinese teachers favoring proactive over reactive instructions, and the U.S. teachers showing the reverse pattern.

8.5. Classroom predictors of behavioral engagement

Second, to further examine whether classroom variables could account for cultural differences in observed behavioral engagement, we used latent growth modeling to account for the nesting of multiple observations within classrooms (Raudenbush & Bryk, 2002). For the 70 intervals (two 30-second intervals per each 35-minute observation) nested within each of 15 classrooms, we used a multi-level modeling program, Hierarchical Linear Modeling (HLM) 6.02 (Raudenbush, Bryk, Cheong, & Congdon, 2005), which estimates regression coefficients with correct standard errors corrected for nesting. Growth modeling also allowed us to disentangle variance of intercept and slope parameters into interval and classroom components, and accounts for each type of variation. In the present study, we controlled for classroom-level variation and investigated associations among country, group size, and types of teacher instruction with student behavioral engagement.

We built a two-level model with observations of student engagement as the outcome variable in the level 1 model, controlling for classroom level variation at level 2. In our level 1 model, we created a latent growth model of the outcome, behavioral engagement (coded 0, 1 at the interval level), with observation at level-1 and classroom at level-2. To maintain parsimony, which is important with a small number of classrooms, we then dropped the variables that did not significantly predict the outcome (Raudenbush & Bryk, 2002).

Level 1

$$Y_{jt} = \pi_{0j} + \pi_{1j}M_{jt} + e_{jt} \quad (1)$$

Level 2

$$\begin{aligned} \pi_{0j} &= \beta_{00} + r_{0j} \\ \pi_{1j} &= \beta_{10} + r_{1j} \end{aligned} \quad (2)$$

As shown in Equation (1), Y_{jt}, the percent of observations coded behaviorally engaged at time t (or, the probability of classroom j being engaged at time t), is a function of the systematic growth trajectory plus random error (e_{jt}) at level-1, with M_{jt} as the linear time code starting with 0 for the first observation and 69 for the last. Therefore, the model intercept represents the percent of engaged behavior at the first observation, and the slope parameter reflects the linear change rate for one 30-second interval. It was assumed that e_{jt} is independently and normally distributed with a mean of zero and constant variance. Equation (2) further specifies π_0 as a function of the mean percent of classrooms in engaged behavior (β_{00}), plus error for individual classroom j (r_{0j}), and (π_1), the mean change over observations (slope) in engagement for all classrooms, plus error for classroom j's slope.

The linear growth model demonstrated significant classroom-level variance in both intercept, χ^2 $(13, N = 15) = 1270.91$, $p < .01$, and slope, χ^2 $(13, N = 15) = 102.29$, $p < .01$. In other words, engagement levels differed by classroom with regard to initial engagement levels, and in the rate of change over time. Based on this model, 57% of the variance in behavioral engagement

Table 3
HLM results modeling student on-task behavior.

Fixed effects	Coefficient	Approx. *t*-ratio	*df*	*p*-value
Intercept, π_{00}	.89***	21.56	11	.00
Culture, β_{01}	.11	−0.36	11	.73
Group size, β_{02}	−.83*	−2.58	11	.03
Reactive instructions, β_{00}	−1.23*	−2.33	11	.04
For growth rate, π_{10}				
Intercept, β_{10}	−.003**	−3.11	13	.01
Culture, β_{11}	.005*	2.50	13	.03
Random effects	**Coefficient**	χ^2	*df*	*p*-value
Intercept, U0	.02***	91.74	11	.00
Growth rate, U1	.00***	69.87	13	.00
Level-1, R	.07			

Note. *$p<.05$; **$p<.01$; ***$p<.001$.

was among classrooms. We proceeded to build the final model (see below) by adding predictors, one at a time, and deleting those not reaching significance. Although outcome variables need not be normally distributed, it is important that residuals from the final model follow a normal distribution; residuals from the final model met this assumption.

Level 1

$$Y_{jt} = \pi_{0j} + \pi_{1j}M_{jt} + e_{jt} \tag{3}$$

Level 2

$$\begin{aligned} \pi_{0j} &= \beta_{00} + \beta_{01}(\text{country}) + \beta_{02}(\text{group size}) + \beta_{03}(\text{reactive instructions}) + r_{0j}; \\ \pi_{1j} &= \beta_{10} + \beta_{20}(\text{country}) + r_{1j} \end{aligned} \tag{4}$$

The final level-1 model, Equation (3), was the same as in Equation (1), but we added three classroom predictors at level-2. The percent of engaged behavior was comprised of the classroom mean at time 1, and included the effect of country, activity setting (calculated as an aggregate of group size across observations), the effect of the classroom percentage of reactive instructions, and random error associated with classroom j (Equation (4)). In initial models, we included the percentage of teacher proactive instructions, but this effect was non-significant, $t\,(10) = -1.30$, $p > .05$. To maintain model parsimony, and because the level-2 *n* was small, we omitted this variable from the final model (Raudenbush & Bryk, 2002). The two continuous variables, group size and reactive instructions were grand-mean-centered. Country was uncentered and dummy coded (China = 0; United States = 1); therefore, the intercept reflects the average behavioral engagement in Chinese classrooms.

Table 3 shows the final results for the HLM analysis. Comparison of the final model, which added classroom predictors, with the linear growth model without predictors demonstrated that the final model explained 77% of the variance in intercept and 50% of the variance in the growth rate. For the Random Effects however, it was evident that significant classroom-level variability remained for the intercept, χ^2 (11, $N = 15$) = 91.74, $p < .01$, and slope, χ^2 (13, $N = 15$) = 69.87, $p < .01$.

Looking at the Fixed Effects (country, group size, and timing of instructions) revealed that overall, as time progressed, engagement decreased significantly, $t\,(11) = -3.11$, $p < .01$, $d = -.15$. There was no main effect of country on engagement, $t(11) = -.36$, *ns*. However, country had a significant effect on behavioral engagement through an interaction with time, $t(11) = 2.50$, $p < .05$. Engagement in American classrooms decreased sharply over time, whereas Chinese behavioral engagement decreased, but comparatively less. Fig. 1 shows behavioral engagement over time for both countries, with the dotted line representing the probability of engagement with each observation in China, and the solid line representing the probability of engagement in the United States.

Controlling for other predictors in the model, group size was significantly associated with engagement, such that intervals coded with students in small groups were less likely to be coded as engaged, $t(11) = -2.58$, $p < .05$, $d = .58$. Finally, classrooms with a high average of reactive instructions had less student behavioral engagement, $t(11) = -2.33$, $p < .05$, $d = .28$, but this did not differ by country (i.e., the interaction between country and reactive instructions was non-significant, $t(11) = -.027$, *ns*, which was dropped from the model.)

The final model and results are summarized in Table 3. To ensure that significant predictors in our model did not predict the outcome because they covaried with each other, we also examined correlations between the significant predictors. Nonparametric correlation analysis (Kendall's *tau_b*) showed there was no correlation between the two significant predictors in the model: activity setting and reactive instructions, $r(13) = -.04$, *ns*, across countries or within a country, $r(5) = -.004$ and $r(6) = -.038$, *ns*. In addition, country was correlated with activity setting, with Chinese culture associated with large group work versus small-group and individual work, $r(13) = .45$, $p < .01$. Country was not associated with reactive instructions, $r\,(13) = .04$, *ns*.

9. Discussion

This study examined behavioral engagement, activity setting, and teacher instructions about behavior during mathematics lessons in China and the United States. Our findings extend prior cross-cultural work that has focused on achievement and curricular differences. Three main findings emerge, demonstrating greater variability in engagement in American classrooms relative to Chinese classrooms, and systematic relations among engagement, group settings, and teacher regulatory instructions.

9.1. Cultural differences in classroom variability

Our study indicated that variability in all observed variables was more pronounced in American than in Chinese classrooms. This corroborates prior research highlighting stark differences in the educational experiences of American students, whereas variability does not characterize schooling in other countries to the same degree (Cameron et al., 2005; Pianta et al., 2002; Stevenson & Stigler, 1992; Tobin, Wu, & Davidson, 1989). Multiple factors may contribute to this difference in variability, including political as well as cultural influences. For example, a central control system in China (e.g., the Administration of Education in China) provides specific and consistent curriculum, goals, and teaching strategies and materials for each classroom in each district. In contrast, American curricular guidelines are generally less specific and authoritative (Cohen & Spillane, 1992). The present sample reflects this diversity in materials selection and lesson plan. For example, the schools in Beijing were provided standardized textbooks and specific class guidelines, whereas the Midwestern schools in this study had the freedom to choose their textbooks and develop their own class plan around the state learning standards. There are likely additional reasons for the variability in the United States such as wide variability in self-regulation skills (Shonkoff & Phillips, 2000).

9.2. Teacher behavioral instructions as an organizer of behavioral engagement

Though distal factors related to curriculum selection and teacher training are important, more proximal features of the classroom including teacher behaviors are strong predictors of students' classroom functioning (Mashburn et al., 2008). For example, research shows behavioral engagement increases in well-organized classrooms where teachers explicitly encourage student self-regulation (Evertson et al., 1983; Pressley et al., 2001). We found that Chinese teachers gave relatively more proactive task-related instructions and American teachers gave relatively more reactive or correctional behavioral instructions. Experts in classroom management have written extensively about how teachers "socialize self-guidance" in students. Teacher behaviors include clarifying expectations and supporting students' abilities to function independently in classrooms by giving directions in advance (Brophy, 1985, 1988; Brophy & Good, 1986). However, before now, little research has examined this point cross-culturally.

Although Chinese teachers gave more than twice as many proactive instructions as their U.S. peers, the number of reactive instructions was roughly the same in the two countries (see Table 1). Interestingly, only reactive instructions significantly predicted behavioral engagement, regardless of country, with more reactive instructions associated with lower behavioral engagement. We note that teacher behavior and student engagement are likely intertwined in meaningful ways. For example, reactive regulatory instructions themselves may be distracting, such as when a teacher scolds one child for looking out the window, and all the children follow by looking out the window. Another possibility is that teacher reactive instructions are an index of (and a response to) student disengagement. That is, when students become less behaviorally engaged, teachers tend to respond with reactive rather than proactive strategies. If so, we might also expect that proactive strategies would be associated with greater behavioral engagement. In our sample however, the majority of proactive instructions happened in the Chinese classrooms, where most of the time, students were engaged. As a result, we might not have had enough variability to find a significant contribution of proactive instructions to student engagement.

9.3. Activity setting and behavioral engagement

Classrooms in both countries were more likely to be coded as engaged in teacher-directed large-group activity settings than in small-group settings. Nonetheless, activity setting was also somewhat culturally embedded; in China, almost all instruction was spent in whole-group settings. Prior work suggests behavioral requirements are less demanding during whole-group time compared to independent work, perhaps because teachers instead of students direct attention and set parameters for appropriate actions (Rimm-Kaufman et al., 2005). In contrast, small-group activities challenge behavioral skills, when students must carry out plans with minimal teacher guidance. Our results substantiate this hypothesis, because classrooms were less likely to be coded as engaged during child-directed small-group settings.

It is possible that during whole-group settings (in either or both countries), students exhibited compliance or passive engagement, which we included as part of behavioral engagement (Fredricks et al., 2004; Wachs, Gurkas, & Kontos, 2004). Notably, the decrease of classrooms coded as behaviorally engaged over time occurred in both countries, regardless of setting, yet most strikingly for American classrooms. This finding highlights two points: First, engaged behavior in both countries declined over time, suggesting fatigue might set in for all students regardless of country. Second, despite fatigue, Chinese students were better able to maintain their attention throughout the lesson, whereas American student's on-task behavior dropped more sharply by time. Research on factors contributing to fatigue and ways teachers maintain student attention throughout a lesson is relatively sparse, but may prove valuable for understanding how to maximize learning. For example, recess is thought to be important for student engagement. Pellegrini and Davis (1993) found that American elementary children became progressively inattentive when recess was delayed, resulting in more active play when recess occurred. American schools have greater variability in recess breaks compared to school systems in other countries, such as British schools where all students have three recess periods of 15 minutes each (Pellegrini & Smith, 1993). Chinese schools, including the ones in this sample, typically have a 10-minute recess after each lesson and a longer recess (often 20 minutes) after every two classes. We argue that having regular recess time might not only help students expend physical energy but also provide clear expectations for when children are expected to focus and when they can take a break.

9.4. The influence of other aspects of teaching and classroom practice on engagement

Other, unmeasured aspects of the classroom and teacher may also directly or indirectly influence student engagement. These include cultural values, teacher experience and beliefs, student beliefs and classroom crowdedness. For example, respect for elders (i.e. teachers) in collective societies like China may lead to more student engagement in the classrooms than in individualistic societies such as the U.S. Additionally, teacher beliefs were found to predict diverse aspects of the classroom practice such as the degree of student autonomy provided by the teacher (Stipek, Givvin, Salmon, & MacGyvers, 2001), and in China, attitudes about mathematics may relate to how instruction is implemented (An, Kulm, & Wu, 2004). Another possibility includes differences in learning conditions across the two countries, such as classroom crowding. The impact of class size or a classroom crowding effect on student engagement is still under debate, but some studies have shown a negative effect of large class size on learning in the United States, (e.g. Krueger, 2002). Some researchers have found that it is peer effects, rather than classroom crowding, that negatively influence children's achievement (Ahmed & Arends-Kuenning, 2006). In other countries, researchers have argued that the effect of more students in a single classroom was positive, using samples from South Africa (Cass & Deaton, 1999) and Israel (Angrist & Lavy, 1999). The Chinese classrooms in our sample certainly had more students than the American classrooms, though we did not measure the total size of the classroom.

It seems quite possible, however, that the effects of class size would be mediated by teaching practices. If larger class sizes predispose teachers toward giving clear, proactive instructions, this might lead to greater rather than lesser student engagement.

Future studies taking into account teacher belief and situational factors can build on the current findings and further explore potential contributors to the differences we observed across cultures. It is worth reiterating that the purpose of this study is to identify potentially valuable educational practices embedded in different countries instead of comparing them with an evaluative point of view. Causal conclusions are not possible based on our study and undoubtedly, other factors are probably associated with student engagement.

10. Implications and future directions

To fully understand effective classroom practices, one nonetheless needs to consider relations among multiple variables, while remembering that many of these practices may be culturally embedded. We consider some of these connections with the following four implications from this study.

First, we highlight the transactional nature of student engagement and teacher efforts. When children are off-task, teachers may be more reactive; they may also have less time or energy to focus on proactive strategies. Reactive instructions may suggest to students that the teacher, not the student him or herself, is responsible for correcting behavior. This is an untested hypothesis that could be explored by asking students to report on their perceptions of classroom management strategies, or by observing the behavior of individual students after being corrected.

In addition, although we found large group activity settings positively associated with on-task behaviors, the practices observed in Chinese classrooms – large-group instruction, for instance – are not necessarily effective by definition. Rather, effective practice is a function of teacher factors including execution of the lesson, as well as student behavior. It is possible that cultural differences in teacher practices contribute to more effective instruction in Asian versus American classrooms. In the current sample, for example, the Chinese teachers tried to promote active engagement consistently during the whole-class as well as small-group time by giving proactive instructions, whereas the U.S teachers tend to use the whole-class setting as primarily lecture time, where reactive instruction occurred more frequently than proactive instructions. Previous research has provided additional evidence. Consistent with this rationale, Wang and Lin (2005) argued that Chinese teachers used increasingly sophisticated mathematics problems in a whole-group setting to engage their students in integrating current and previously learned concepts and providing justifications for their problem solutions. Based on our anecdotal observations from this study, Chinese teachers moved the lessons from stages of instruction, to guided practice, and then to independent practices. In contrast, American mathematics teachers have been found to cover many topics in an unsystematic and non-intuitive manner (Stevenson, Lee, & Stigler, 1986). To understand these relations, studies might examine the coherence of a lesson and the amount of active participation and discussion (i.e., behavioral engagement).

Second, this study informs professional development by revealing the type of teacher instructions (i.e., reactive) that occur with off-task behavior. Reactive instructions were negatively associated with student engagement regardless of country and classroom setting. On the one hand, studying what is disruptive about reactive instructions might be enlightening;

on the other hand, exploring cross-cultural differences in what engages students in the classroom may prove most helpful. For example, prior research has suggested that proactive instructions may help promote behavioral engagement by providing a preview of expectations (Bohn et al., 2004; Cameron et al., 2005). These studies also suggest that teacher actions early in the school year, and early in the school day, help shape the classroom environment and student reactions to the classroom.

Uncovering ways to promote behavioral engagement, such as encouraging teachers to maximize proactive behavioral instructions, minimize reactive instructions, and maximize opportunities for students to be actively engaged, may contribute to higher levels of on-task behavior and perhaps, subsequent learning. Thus, a third implication for professional development is to ask novice teachers to observe videotapes of experienced teachers who are more proactive in teaching and using strategies such as self-regulatory instructions to keep students on task. Asking teachers to comment on videos from a different country would help us understand what activities or instructions teachers value as individuals and as individuals within a country. In addition, studies addressing the influence of teacher experience, beliefs and student attitudes towards effective classroom management would not only help us disentangle their effects from teacher regulating behavior, but also provide important information for teacher training. Again we note that activities in one country may have different implications in the context of country.

Finally, this study focuses on first grade elementary school classes because during this period children first start learning to regulate themselves in well-organized math classes. Understanding various classroom and teacher factors is crucial to understanding children's engagement and associated math learning. The results should also apply to learning in general in other age groups such as preschool and kindergarten, in which teachers spend a substantial amount of time regulating students and organizing classrooms. Studies incorporating a more comprehensive list of classroom and teacher factors would be helpful.

11. Limitations

Despite the new information gleaned from this study, three limitations must be mentioned. We employed a strategy that allowed us to investigate only the first 35 minutes in each classroom. Because the Chinese lessons were in general 4 minutes longer than the U.S. lessons, we acknowledge the possibility of losing interesting information at the end of the Chinese lessons. However, comparing engagement in a similar period of time is informative to see what happens in the same amount of time in two different settings. Interestingly, engagement levels dropped gradually rather than at the end of the lessons. In particular, the engagement of the American students started dropping several minutes after the beginning of the lesson. Our data did not indicate a sudden engagement drop in the last 5 minutes in China. Future studies comparing longer American classes with full length Asian classes would further address this issue. Another limitation related to coding is that some information might be lost due to our decision to dichotomize classrooms on engagement (if more than 70% engaged was coded as on task) based on previous research. Future research should extend this study by exploring student engagement at an individual level.

Second, although videotape data were rich and yield significant results, our sample included only one day's observation from a small number of classrooms. Video data from multiple days and classrooms should be collected across varied lessons and activities in order to obtain a representative picture of classroom functioning.

Third, because engagement was not assessed for individual children outside the classroom context, it is difficult to disentangle how culture might contribute to individual levels of behavioral engagement from effects of activity setting. That is, Chinese students may have been more often engaged because they spent relatively more time in whole-group settings, which have been said to pose less behavioral self-regulation demands. It is also possible that compared to American children, Chinese students had better self-regulatory skills, which made it easier for them to be engaged (Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Including multiple indicators of engagement in future research, including at the student level, is vital for understanding the implications of culture, activity setting, and teacher instructions for individual student outcomes, including achievement. Distinguishing among these may tell us much about the types of engagement likely to happen in different settings, and importantly, specific practices that teachers of large classes use to promote active engagement. Finally, although we carefully matched schools in the two sites by school reputation and general family background, it would be helpful to collect detailed background information from each child in our study.

12. Summary and conclusion

This study used a cross-cultural lens to examine relations among classroom behavioral engagement during math lessons in the United States and China. Activity settings and teacher task-related instructions each uniquely predicted behavioral engagement in the classroom. Moreover, there were cultural differences, such that students in Chinese classrooms were more often engaged than U.S. students, even in whole-group activity settings. This suggests other possible contributors beyond curricular and instructional differences to the Asian-American achievement gap in mathematics. Our findings set the stage for further exploration of differences in contributors to student engagement in the classroom as well as classroom practices. Additional studies are needed to understand cultural specificity as well as general, pan-cultural principles of teaching and learning in the school context and their implications for achievement in mathematics.

References

Ahmed, A. U., & Arends-Kuenning, M. A. (2006). Do crowded classrooms crowd out learning? Evidence from the food for education program in Bangladesh. *World Development*, *34*, 665–684.

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, *36*, 369–387.

An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school mathematics teachers in China and the US. *Journal of Mathematics Teacher Education*, *7*, 145–172.

Anderson, L. M., Evertson, C. M., & Emmer, E. T. (1980). Dimensions in classroom management derived from recent research. *Journal of Curriculum Studies*, *12*, 343–362.

Angrist, J., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, *114*, 533–575.

Beaton, A. E., Martin, M., Mullis, L., Gonzalez, E., Smith, T., & Kelly, D. (1996). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*, 647–663.

Blatchford, P., Moriarty, V., Edmonds, S., & Martin, C. (2002). Relationships between class size and teaching: A multimethod analysis of English infant schools. *American Educational Research Journal*, *39*, 101–132.

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

Brenner, M. E., Herman, S., Ho, H. Z., & Zimmer, J. M. (1999). Cross-national comparison of representational competence. *Journal for Research in Mathematics Education*, *30*, 541–557.

Brophy, J. E. (1985). Classroom management as instruction: Socializing self-guidance in students. *Theory into Practice*, *24*, 233–240.

Brophy, J. E. (1988). Educating teachers about managing classrooms and students. *Teaching & Teacher Education*, *4*, 1–18.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 340–370). New York: Macmillan.

Cameron, C. E., Connor, C. M., & Morrison, F. J. (2005). Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology*, *43*, 61–85.

Carta, J. J. (1991). Education for young children in inner-city classrooms. *American Behavioral Scientist*, *34*, 440–453.

Case, A., & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*, *114*, 1047–1084.

Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian American, Caucasian-American, and East Asian high school students. *Child Development*, *66*, 1215–1234.

Clarke-Stewart, A., Lee, Y., Allhusen, V. D., Kim, M. S., & McDowell, D. J. (2006). Observed differences between early childhood programs in the U.S. and Korea: Reflections of 'developmentally appropriate practices' in two cultural contexts. *Journal of Applied Developmental Psychology*, *27*, 427–443.

Cohen, D. K., & Spillane, J. P. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education*, *18*, 3–49.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., & Cooney, T. J. (1985). *Second international mathematics study summary report for the United States*. Champaign, IL: Stipes.

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American Educational Research Journal*, *17*, 43–60.

Evertson, C. M., Emmer, E. T., Sanford, J. P., & Clements, B. S. (1983). Improving classroom management: An experiment in elementary school classrooms. *Elementary School Journal*, *84*, 173–188.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*, 59–109.

Geary, D. C., Bow-Thomas, C. C., Fan, L., & Siegler, R. S. (1993). Even before formal instruction. Chinese children outperform American children in mental addition. *Cognitive Development*, *8*, 517–529.

Greenwood, C. R. (1991). Longitudinal analysis of time, engagement, and achievement in at-risk versus non-risk students. *Exceptional Children*, *57*, 521–535.

Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*, *31*, 328–349.

Gump, P. V. (1969). Intra-setting analysis: The third-grade classroom as a special but instructive case. In E. P. Willems & H. L. Raush (Eds.), *Naturalistic Viewpoints in Psychological Research* (pp. 200–220). New York: Holt, Rinehart, & Winston.

Hannum, E., & Wang, M. (2006). Geography and educational inequality in China. *China Economic Review*, *17*, 253–265.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, *24*, 411–482.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *Elementary School Journal*, *101*, 3–20.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Krueger, A. (2002). *Economic considerations and class size*. (NBER Working Paper No. 8875). Cambridge, MA: National Bureau of Economic Research.

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, *70*, 1373–1400.

Linn, M. C., Lewis, C., Tsuchida, I., & Songer, N. B. (2000). Beyond fourth-grade science: Why do U.S. and Japanese students diverge? *Educational Researcher*, *29*, 4–14.

Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, NJ: Lawrence Erlbaum.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., & Bryant, D. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language and social skills. *Child Development*, *79*, 732–749.

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, *21*, 471–490.

Miller, K. F., Kelly, M. K., & Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of mathematical development. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 163–178). New York: Psychology Press.

Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Child Development*, *2*, 279–305.

Miura, I. T., Chungsoon, K. C., Chang, C. M., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of numbers: Cross-national comparisons. *Child Development*, *59*, 1445–1450.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics. Reston*, VA: Author.

National Institute of Child Health and Human Development Early Child Care and the Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal*, *102*, 367–387.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently and why*. New York: Free Press.

Peak, L., Caldwell, N., Owen, E., Stevenson, H., Suter, L., Frase, M., et al. (1996). *Pursuing excellence: A study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Pellegrini, A. D., & Davis, P. L. (1993). Relations between children's playground and classroom behavior. *British Journal of Educational Psychology*, *63*, 88–95.

Pellegrini, A. D., & Smith, P. K. (1993). School recess: Implications for education and development. *Review of Educational Research*, *63*, 51–67.
Perry, M. (2000). Explanations of mathematical concepts in Japanese, Chinese, and U.S. first-and fifth-grade classrooms. *Cognition and Instruction*, *18*, 181–207.
Peterson, P. L., & Fennema, E. (1985). Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. *American Educational Research Journal*, *22*, 309–335.
Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*, 144–159.
Pianta, R. C., La Paro, K. M., Payne, C., Cox, M. J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *Elementary School Journal*, *102*, 225–238.
Ponitz, C. C., Rimm-Kaufman, S. E., Curby, T. R., & Grimm, K. J. (in press). Kindergarten classroom quality, behavioral engagement, and reading achievement. *School Psychology Review*.
Pressley, M., Wharton-McDonald, R., Allington, R., Block, C. C., Morrow, L., Tracey, D., et al. (2001). A study of effective first grade literacy instruction. *Scientific Studies of Reading*, *5*, 35–58.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Raudenbush, S. W., Bryk, A., Cheong, Y. F., & Congdon, R. (2005). *HLM6: Hierarchical Linear and Nonlinear Modelling*. Lincolnwood, IL: Scientific Software International.
Rimm-Kaufman, S. E., La Paro, K. M., Downer, J. T., & Pianta, R. C. (2005). The contribution of classroom setting and quality of instruction to children's behavior in the kindergarten classroom. *Elementary School Journal*, *105*, 377–394.
Robitaille, D. F., & Garden, R. A. (Eds.). (1989). *The IEA Study of Mathematics II: Contexts and outcomes of school mathematics*. New York: Pergamon Press.
Romberg, T. A. (1997). The influence of programs from other countries on the school mathematics reform curricula in the United States. *American Journal of Education*, *106*, 127–147.
Romberg, T. A. (1999). School mathematics: The impact of international comparisons on national policy. In G. Kaiser, E. Luna, & L. Huntley (Eds.), *International comparison in mathematics education* (pp. 189–199). Philadelphia, PA: Falmer Press.
Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*, 74–81.
Sanford, J. P., & Evertson, C. M. (1983). Time use and activities in junior high classes. *Journal of Educational Research*, *76*, 140–147.
Santagata, R., & Stigler, J. W. (2000). Teaching mathematics: Italian lessons from a cross-cultural perspective. *Mathematical Thinking and Learning*, *2*, 191–208.
Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the consequences*. Boston: Kluwer Academic Publishers.
Schleppenbach, M., Perry, M., Miller, K. F., Sims, L., & Fang, G. (2007). The answer is only the beginning: Extended discourse in Chinese and U.S. mathematics classrooms. *Journal of Educational Psychology*, *99*, 380–396.
Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC, US: National Academy Press.
Stallings, J. (1975). Implementation and child effects of teaching practices in follow through classrooms. *Monographs of the Society for Research in Child Development*, *40*, 1–133.
Stevenson, H. W., & Lee, S. (1995). The East Asian version of whole class teaching. *Educational Policy*, *9*, 152–168.
Stevenson, H. W., Lee, S., Chen, C., & Stilger, J. W. (1990). Contexts of achievement: A study of American. Chinese and Japanese children. *Monographs of the Society for Research in Child Development*, *55*, 123–1123.
Stevenson, H. W., Lee, S., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science*, *231*, 693–699.
Stevenson, H. W., Lee, S., Chen, C., & Lummis, M. (1990). Mathematics achievement of children in China and the United States. *Child Development*, *61*, 1053–1066.
Stevenson, H. W., & Stigler, J. W. (1992). *The Learning Gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Simon & Schuster.
Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, *35*, 87–100.
Stigler, J. W., & Hiebert, J. (1999). *The teaching gap*. New York: Free Press.
Stigler, J. W., & Perry, M. (1988). Mathematics learning in Japanese. Chinese and American classrooms. *New Directions for Child Development*, *41*, 41.
Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*, 213–226.
Tobin, J. J., Wu, Y. H., & Davidson, D. H. (1989). *Preschool in three cultures: Japan, China, and the United States*. New Haven, CT: Yale University Press.
Wachs, T. D., Gurkas, P., & Kontos, S. (2004). Predictors of preschool children's compliance behavior in early childhood classroom settings. *Journal of Applied Developmental Psychology*, *25*, 439–457.
Wang, J., & Lin, E. (2005). Comparative studies on U.S. and Chinese mathematics learning and the implications for standards-based mathematics teaching reform. *Educational Researcher*, *34*, 3–13.
Yang, M. T. L., & Cobb, P. (1995). A cross-cultural investigation into the development of place-value concepts of children in Taiwan and the United States. *Educational Studies in Mathematics*, *28*, 1–33.
Zanolli, K., Daggett, J., & Pestine, H. (1995). The influence of the pace of teacher attention on preschool children's engagement. *Behavior Modification*, *19*, 339–356.