Leadership, learning-centered school conditions, and mathematics achievement: What can the United States learn from top performers on TIMSS?

Nianbo Dong and Xiu Chen Cravens
Vanderbilt University, Nashville, Tennessee, United States

Drawing on crossnational datasets, including contextual questionnaires and the mathematics achievement results of eighth graders, from the Trends in International Mathematics and Science Study of 2007 (TIMSS 2007), this study examines the viability of using the findings of international assessment reports to inform school leadership practice directed at enhancing learning conditions. The study first identifies core school conditions within the realm of influence that can be captured by the TIMSS conceptual framework, and then examines the association between these conditions and student achievement in mathematics. We focus on education systems whose eighth graders consistently gain higher average scores than their counterparts in the United States on mathematics assessments. These systems are Chinese Taipei, Korea, Singapore, Hong Kong SAR, and Japan. Analyses of the 2007 contextual survey results from the five education systems and the United States reveal interesting differences in school-level conditions for learning and in the associations between these conditions and mathematics achievement. Our initial crossnational analyses indicate links between achievement and learning conditions such as evaluation of the instruction curriculum and instructional implementation, and learning culture. However, the types and strengths of the associations appear to vary according to national context. Findings also indicate discrepancies between the perspectives of teachers and principals regarding school learning conditions.

IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments Volume 5

INTRODUCTION

Around the world, governments are paying increasing attention to the results of large-scale crossnational assessments of student achievement, and are using the findings to inform educational reform initiatives. In the United States, studies of student performance on international assessments are playing an ever more important role in setting reform agendas at both state and national levels (Swanson & Barlage, 2006). In 2008, the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), and Achieve Inc. formed an advisory group to develop "international benchmarking" as a critical tool for creating a world-class education system for United States students. *Benchmarking for Success*, the report issued by the advisory group (NGA, 2008), cautioned that the United States "is falling behind other countries in the resource that matters most in the new global economy: human capital" (p. 5). Drawing together results from international assessments conducted during recent decades, the report called on both state and federal policymakers to provide stronger support for research and development in order to identify and learn from "top performers and rapid improvers," and thereby gain insights and ideas unlikely to be "garnered solely from looking within and across state lines" (p. 6).

Recent large-scale international assessments, such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA), reveal United States students lagging behind their peers in other systems, especially those in East Asia, with respect to mathematics and science achievement (Ferraro & Van de Kerckhove, 2006; Mullis, Martin, & Foy, 2008; Provasnik, Gonzales, & Miller, 2009). For mathematics, in particular, the TIMSS 2007 results found that while the United States fourth-graders and eighth-graders scored, on average, above the international-scale and the US TIMSS 1995 averages, the performance of both cohorts was below that of their peers in several other systems. The fourth graders were outperformed by peers in eight of the 36 participating systems (Hong Kong SAR, Singapore, Chinese Taipei, Japan, Kazakhstan, Russian Federation, England, and Latvia) and the eighth graders by five of the 48 participating systems (Chinese Taipei, Korea, Singapore, Hong Kong SAR, and Japan; Provasnik et al., 2009).

The increasing recognition of the relevance that international assessments have for school improvement initiatives has intensified research interest in analyzing and gaining insight from the crossnational data now made publically available by the testing organizations (Rutkowski, Gonzalez, Joncas, & von Davier, 2010). In order to offer educational stakeholders a fuller appreciation of what the achievement results mean and how they can be used to improve student learning, international assessments such as TIMSS and PISA also incorporate contextual questionnaires designed within the context of theoretical frameworks that aim to capture attributes associated with student learning and achievement (Mullis et al., 2008).

The developing emphasis on studying the major characteristics of educational and social contexts with a view to improving student learning has led to crossnational assessments gradually moving away over recent years from simple descriptions of contextual differences toward identification of factors that are malleable—changeable—and therefore meaningful to those involved in the process. In short, new research questions are being raised that differ from those asked and explored in much of the existing body of work (Baker, Goesling, & LeTendre, 2002; Chudgar & Luschei, 2009; Heyneman & Loxley, 1983).

Today, researchers want to know *what* makes a difference in student achievement in terms of input into schooling. Is it, for example, national income, school characteristics such as size and resources, or student socioeconomic background? They also want to explore how the schooling process affects learning outcomes. The purpose of the international benchmarking is seen as that of informing the development of domestic interventions—programs, practices, policies—that positively influence education outcomes. As such, the between- and within-country variations in achievement results and other factors arising out of the international assessments provide a rich opportunity for researchers engaged in comparative and international education studies to explore plausible associations between learning conditions and student performances (Baker, Lee, & Heyneman, 2003; Porter & Gamoran, 2002; Schmidt, Rotberg, & Siegel, 2003). Learning conditions define the contexts within which student learning takes place. More specifically, they are the factors that affect students' learning, such as national curriculum standards, resource allocation schemes for schools, classroom instructional approaches, teacher qualifications and professional development, student attitudes, and home support for learning.

The contextual questionnaires of the international assessments also afford opportunities for researchers to examine learning conditions from multiple levels and angles. At the national policy level, studies have examined curricular goals of education systems and how the systems were organized to attain those goals (Baker et al., 2003; Schmidt et al., 2003). Included in this level are, for example, curriculum standards, the rigor and coherency of textbook content, and characteristics of the teaching force (Akiba, LeTendre, & Scribner, 2007; Wang, 2004). At the school level, studies have looked into school characteristics in terms of student composition and resourcing, classroom activities, and pedagogical practices (Clarke et al., 2007; Wang & Lin, 2005). At the student level, research has explored the role that home support, parental involvement, and student attitudes toward learning play in academic achievement (Paik, 2004; Shen, 2005; Wang, 2004).

Conspicuously missing from this line of international assessment literature is the connection between school-level leadership and conditions of student learning. Extensive research on school leadership and educational outcomes identifies school principals as the keystone of successful educational reform; they are critical with respect to the successful interpretation and implementation of and support for improvement interventions (Elmore, 2000; Leithwood, Louis, Anderson, & Wahlstrom, 2004; Rice & Islas, 2001). However, because school leaders impact core learning conditions indirectly and holistically (Leithwood et al., 2004; Louis, Leithwood, Wahlstrom, & Anderson, 2010; Waters, Marzano, & McNulty, 2003), constructing a theoretical framework to capture empirical evidence for the linkage between school leadership and student learning is challenging within any learning context, let alone cross-nationally.

As a reaction to the less than optimal performance of United States fourth and eighth graders in TIMSS 1999, the National Association of Secondary School Principals (NASSP) called for stronger instructor leadership on mathematics and science performance (Rice & Islas, 2001). The 2008 NGA report also underscored the importance of developing school leaders and holding principals accountable for ensuring instructional quality by learning from "international best practices" (p. 28). However, despite the strong rhetoric for action, existing international assessment research has yet to provide evidence that contextual questionnaires offer substantive insight into the school-level practices that are positively associated with student achievement.

In this paper, we examine high-performing education systems through the prism of TIMSS in search of valuable lessons for the United States. We focus on the learning conditions that are likely to be within the realm of influence of school leadership. Specifically, we focus on education systems that have consistently produced higher TIMSS average scores than those of the United States in mathematics—Chinese Taipei, Korea, Singapore, Hong Kong SAR, and Japan.

To present our study, we first provide a twofold background review of (a) international assessment research addressing learning conditions and educational outcomes, and (b) leadership research that identifies core components of learning conditions malleable at the school level by principals and their leadership teams. We then examine the TIMSS contextual questionnaires for the extent to which they capture the core learning conditions. We additionally explore, both theoretically and empirically, the possible associations between core components of learning-centered school conditions and student achievement outcomes, by using TIMSS datasets from 2007 that include contextual-questionnaire and mathematics-assessment results for the participating eighth graders. We ask the following research questions:

1. How do core components of learning-centered school conditions compare among the high-performing systems and with the United States as described by the TIMSS contextual questionnaires?
2. Are core components of learning-centered school conditions associated with student achievement in mathematics when contextual factors are held constant? If yes, how and to what extent?
3. Are the relationships between the key malleable variables and student achievement consistent across the six systems when within-system contextual factors such as curriculum standards, attitudes toward learning, school size, teacher qualifications, and student socioeconomic status are held constant?

BACKGROUND

International Assessments and Learning Conditions

Research on learning conditions can be categorized into two main dimensions: school resourcing (personnel selection, qualifications, and training) and the schooling process (curriculum content, teaching pedagogy, teacher collaborations, and classroom structure). In recent decades, crossnational studies utilizing international assessment results to connect these two dimensions of learning conditions with student achievement have been conducted mainly at the national and aggregate levels. The learning conditions that have attracted the most attention in these studies appear to comprise four areas: (a) curriculum standards, rigor, and coherence, (b) teacher qualifications, (c) pedagogical strategies, and (d) home support for learning.

From their analysis of data from the first TIMSS dataset (1995), Schmidt and his colleagues found strong relationships between curricular content and learning outcomes both across countries and across classrooms within countries, especially in the United States (Schmidt et al., 2003). They suggested that much of the poor performance of the United States students could be attributed to a poorly constructed curriculum. Specifically, from their comparison of the curriculum standards of the best-performing nations, Schmidt and his team identified three essential characteristics that the United States standards were lacking (Schmidt et al., 2003):

- *Focus:* covering a smaller number of topics in greater depth at every grade level, enabling teachers to spend more time on each topic so that all students learn it well before they advance to more difficult content;
- *Rigor:* demanding more advanced learning in subjects such as algebra and geometry; and
- *Coherence:* laying out an orderly progression of topics that follow the logic of the discipline, allowing thorough and deep coverage of content.

Although subsequent studies using TIMSS data yielded less convincing associations between curriculum foci and cross-national achievement variation (Baker et al., 2003; LeTendre, Baker, Wiseman, Boe, & Goesling, 2002), the Schmidt study of early TIMSS results served as a sounding alarm that drew national attention to the weakness of having a splintered curriculum and was widely cited as evidence for the necessity of developing coherent and consistent curriculum standards.

Improving teacher quality is another educational reform priority for the United States and other nations. Using TIMSS 2003 data from 46 systems, Akiba et al. (2007) tested the assumption that teacher quality, measured by certification rate, mathematics major, and teaching experience, is associated with student achievement. The authors also examined the association between access to qualified teachers and the socioeconomic status (SES)-based achievement gap. The authors found from their analyses that the achievement gap in the United States between high-SES and low-SES students was among the largest when compared with the relevant data from other nations that participated in TIMSS 2003. However, they also found that the gap in access to qualified teachers was not significantly associated with the achievement gap between students with high and low SES. Their research left room for further investigation into other influential factors, such as instructional resources and teacher learning opportunities, that might help account for the student achievement variation unexplained by the teacher qualification measures alone.

Analyses of teacher questionnaire items regarding classroom activities and video archives provided data for deeper probes into instructional practices and pedagogical strategies. Givvin, Hiebert, Jacobs, Hollingsworth, and Gallimore (2005) conducted an ethnographic study that drew on the TIMSS 1999 video archives. They used three coding dimensions when analyzing the data—purpose of activity, interaction structure, and content activity. Their findings suggested that within the seven systems (Australia, the Czech Republic, Hong Kong SAR, Japan, the Netherlands, Switzerland, and the United States) that participated in the TIMSS 1999 Video Study, eighth-grade mathematics teachers within a country taught lessons in relatively similar ways. They also found that many of the features within the three dimensions examined were discernible in all seven systems (Givvin et al., 2005).

These findings supported the suggestions made by others (e.g., LeTendre, Baker, Akiba, Goesling, & Wiseman, 2001) that countries share patterns of teaching practice. Such convergence, according to Givvin and colleagues, provides opportunities for educators not only to share familiar notions regarding classroom practices but also to realize that seeing "the familiar in a new light might offer many opportunities for teachers to rethink taken-for-granted practices and to see them as choices rather than inevitabilities" (2005, p. 342). This view was expanded by the research of David Clarke and colleagues, which emphasized the complexity of international and comparative learning associated with instructional practices (Clarke et al., 2007). In differentiating the choices of instructional unit for analysis, they revealed significant structural variation in any one teacher's lesson sequence, suggesting that a single lesson pattern is unlikely to be an accurate or a useful representation of either an individual teacher's lessons or of any nationally representative sample of lessons. Furthermore, Clarke et al. (2007) raised important questions about the assumption that less-successful countries would necessarily do well to adapt instructional practices of countries consistently successful on international measures of mathematics performance, given that variations in student performance might be attributable to other differences in culture, societal affluence, or aspiration.

The search for other factors influencing student learning beyond and/or in connection with classroom instructions has led researchers to explore the role of home support, parental involvement, and student attitudes toward learning in academic achievement (Paik, 2004; Shen, 2005; Wang, 2004). For example, using TIMSS 1995 data, Wang (2004) compared the mathematics achievement of students from Hong Kong SAR with that of their peers from the United States. Wang also looked at a series of family background factors, such as mothers' expectations, presence of study aids, and extracurricular time spent on various activities. Finding that some of the factors differentially influenced the Hong Kong and the United States students, Wang conjectured that the differences could be culture-dependent. Paik (2004) used a multiple-factor psychological model to analyze the home and school factors in TIMSS 1995 for Korean and United States students. The findings of the study suggested implications for family–school partnerships, after-school or weekend programs targeted at improving academic competencies, and family support, given that all appeared to be influential factors for learning.

Shen (2005) used multivariate discriminant analysis and data from TIMSS 1999 to make comparisons between the United States middle school system and five top-ranked Asian middle school systems, with respect to student achievement in mathematics and science. The analyses were based on variables related to school and classroom environment as well as students' out-of-school life, home background, and self-perceptions about mathematics and science ability. The results further illuminated the differences between American schools and their Asian counterparts, especially the somewhat more peripheral place of schooling in the lives of American adolescents versus the more central position of schooling in the lives of their East Asian peers.

Malleable Learning Conditions and School Leadership

Despite the complexity and difficulty of identifying factors and conditions that can be optimized for student learning, research suggests that strategic actions which integrate core components of school-wide improvement efforts are essential in terms of effective reform efforts for student learning (Desimone, 2006; Goldring, Porter, Murphy, Elliott, & Cravens, 2009; Rowan, Correnti, Miller, & Camburn, 2009). Furthermore, a relatively extensive research base supports the notion that school leaders play a pivotal role in interpreting, implementing, and sustaining such intervention measures by enhancing teaching and learning conditions (Hallinger & Heck, 2010; Leithwood et al., 2004; Waters et al., 2003). But why is leadership crucial? Louis and colleagues conjectured, on the basis of their six-year leadership study, that "leaders have the potential to unleash latent capacities in organizations" (Louis et al., 2010, p. 7). They pointed out that while most school variables, considered separately, have only small effects on student learning, it is possible to create synergy across the relevant variables operating among the key stakeholders in the process. Educators in leadership positions, as Louis and her colleagues asserted, are uniquely well positioned to ensure the creation and sustainability of this synergy.

Intervention-oriented research, therefore, focuses on understanding the nature of strong leadership and, more specifically, of identifying the pathways through which leadership affects learning conditions. Research in recent decades suggests that school leaders impact student learning by establishing school conditions which support and strengthen teaching and learning (Waters et al., 2003). Much of the evidence shows, however, that the connection is indirect and complex. Leadership studies suggest that the direct and indirect effects of school leadership on student learning are small but significant at about five to seven percent of the variation in student learning across schools, or about one quarter of the total across-school variation (12 to 20%) explained by all school-level factors after controlling for student characteristics (Creemers & Reezigt, 1996; Louis et al., 2010). Further empirical evidence suggests that school principals, along with their leadership teams, influence student outcomes by enhancing curriculum structures and instruction practices as well as providing academic support for parents and students (Cohen & Hill, 2000; Goldring & Cravens, 2007; Leithwood & Jantzi, 1999; Smith, Desimone, & Ueno, 2005).

In preparation for reauthorization of the Elementary and Secondary Education Act, the U.S. Department of Education issued *A Blueprint for Reform*, which called on "states and districts to develop and implement systems of teacher and principal evaluation and support, and to identify effective and highly effective teachers and principals on the bases of student growth and other factors" (U.S. Department of Education, 2010, p. 4). While the term "effective" in conjunction with leadership or teaching is often treated with considerable skepticism, a comprehensive review of the research literature (Goldring et al., 2009; Porter, Goldring, Muphy, Elliott, & Cravens, 2006) reveals six core components of leadership that are highly effective with regard to student learning and achievement: holding high standards for student performance, a rigorous curriculum, quality instruction, a culture of learning and professional behavior, connections to external communities, and systemic performance accountability.

The learning-centered leadership framework has three strong features that provide the scaffold for empirical research designed to detect malleable school conditions for learning. First, the focus of this framework is on measureable leadership behaviors drawn from literature on effective schools and school districts. This framework fits within a more general leadership model (CCSSO, 1996; Glasman & Heck, 1992; Hallinger & Heck, 1996) of what qualifications school principals must have and how principals in the school system are expected to perform, but it does not try to address every aspect of the overall leadership process. The framework focuses on principal behaviors that are linked to teachers' opportunities to improve instructional practices. Not included in the framework are other aspects of leadership that are considered to be the precursors of leadership behaviors, such as knowledge and skills, personal characteristics, and beliefs (Murphy, Goldring, Elliott, & Porter, 2006). Second, the core components include standards, curriculum, instruction, culture, external environment, and performance accountability, which rest upon the same theoretical foundation as that of the international-assessment contextual frameworks (Mullis, Martin, Smith, Garden, Gregory, & Gonzalez, 2005). Third, the learning-centered core components assume that there are aspects of the context within which leadership and schooling take place that might moderate the impact of leadership effects. For example, everything else being equal, the evaluation of leadership quality might appropriately take into account systemic curriculum standards, experience of leadership, length of time in the same school, student body composition, staff composition, level of schooling, and the geographic setting of the school.

In summary, our review of literature indicates that, to date, while a substantial number of studies have been conducted using large-scale international assessment data (e.g., Rutkowski et al., 2010) and while some have explored the connection between school contexts and student achievement, few have drawn on the available cross-national datasets to address the role of school leaders in improving learning conditions. Even fewer have applied theoretically-grounded frameworks and sophisticated analytical methodology.

METHOD

Data

TIMSS 2007 was the fourth administration (since 1995) of international benchmarking of Grade 4 and Grade 8 student achievement in mathematics and science undertaken by the International Association for the Evaluation of Educational Achievement (IEA). In 2007, 36 educational jurisdictions participated in the Grade 4 testing and 48 participated in the Grade 8 testing. Participating systems administered the TIMSS assessments to two system-wide probability samples of schools and their students, based on a standardized definition. Countries were required to draw samples of students who were nearing the end of their fourth year or eighth year of formal schooling. The sample included both public and private schools, randomly selected and weighted to be representative of the nation. Achievement results from TIMSS are reported on a scale that has a scale average of 500 and a standard deviation of 100 (Gonzales et al., 2008).

Although assessment results were available for both Grades 4 and 8 for this current study, we decided to focus on Grade 8 for several reasons. Middle school grades are considered critical to formative adolescent development and cognitive learning yet are also considered to be, within the context of school management, the grades most susceptible to adverse academic and social factors (Cobb & Smith, 2005). Furthermore, because elementary and secondary schools in most education systems have separate and often different administrative structures, focusing on learning conditions at the secondary school level (here, Grade 8) may add to the clarity of the findings.

Six education systems featured in our crossnational study: Chinese Taipei, Korea, Singapore, Hong Kong SAR, Japan, and the United States. Our choice of the first five systems was based on the fact that they had statistically higher overall average mathematics scores than the other systems, including the United States, participating in TIMSS in 2007 (see Table 1). The same five education systems also had consistently higher average mathematics scores than the United States in TIMSS 1999 and TIMSS 2003.

The dependent variable for the analyses was five plausible values of mathematics score for each student from TIMSS 2007. The independent variables were derived from the accompanying contextual questionnaires administered to the school principals, teachers, and students sampled in each system in 2007 and will be further discussed in this paper as measures for learning-centered and malleable school conditions.

Table 1: Weighted mean national mathematics achievement scores on TIMSS 2007

Education system	Mean	School *N*	Student *N*
Chinese Taipei	597	143	3,830
Korea	597	144	4,072
Singapore	592	155	4,351
Japan	571	141	4,151
Hong Kong SAR	569	106	3,040
United States	508	192	5,859

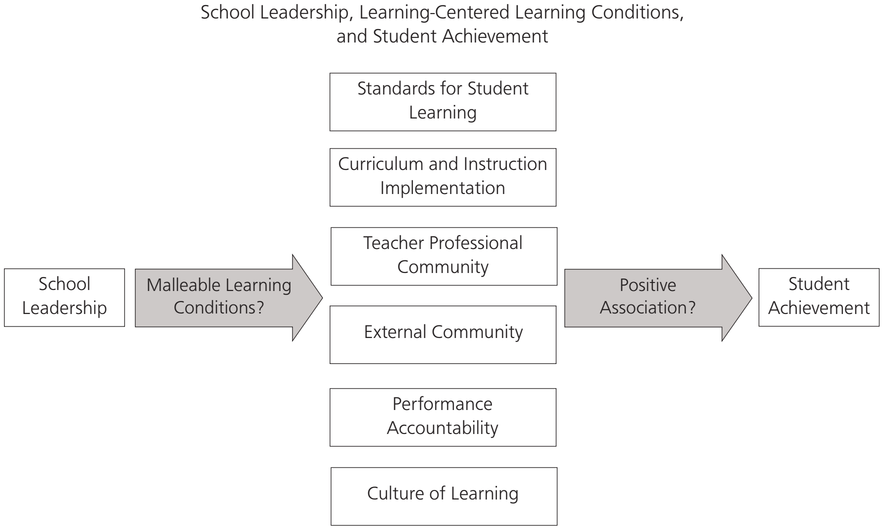
Source: TIMSS 2007 Mathematics Grade 8 database.

We handled missing data by first using listwise deletion for the observations with missing data in all key variables. We then applied the multiple imputation technique to impute the missing data. For the principal and teacher surveys, we imputed once. For student data, we imputed once based on each of five plausible values of the student mathematics score. We next merged principal, teacher, and student data, a process that gave us five imputed datasets for final analysis. The final dataset included a sample of 25,303 students nested in 881 schools of six education systems.

Our construction of measures for learning-centered school conditions using TIMSS 2007 contextual questionnaires was a three-step process. First, we studied the contextual framework document issued by IEA (Mullis et al., 2005) for intended construct domains to be covered by the school, teacher, and student questionnaires. Second, we combed through the questionnaires and identified items that met the face validity criteria for measuring dimensions of learning-centered school conditions. That is, we compared the items listed in the TIMSS conceptual framework as covering school learning conditions with the descriptions provided by the learning-centered leadership framework, and thereby formed the initial variables for the study. Third, based on preliminary exploratory factor analysis and the theory that we reviewed, we grouped selected variables into an analytic framework containing six measures of learning-centered school conditions: (a) standards for student learning, (b) curriculum and instruction implementation, (c) teacher professional community, (d) external community, (e) performance accountability, and (f) culture of learning (see Figure 1).

Our analytic framework essentially modified the core components of the Goldring et al. (2009) learning-centered leadership model, based on the literature review of international comparisons of student performance. We treated *culture of learning* as a stand-alone component in order to capture the importance of learning motivation in each national context. This consideration is especially relevant to the fact that all five systems with higher TIMSS average mathematics scores than those of the United States are East Asian and, as suggested in the literature, share a strong meritocratic cultural background that values education (Paik, 2004; Shen, 2005). We decided to merge and capture the core components of curriculum rigor and instructional quality with *curriculum and instruction implementation*, given that the available items in the TIMSS school and teacher questionnaires tended to address the two dimensions in a combined manner. Specific TIMSS questionnaire items and how they contributed

Figure 1: Conceptual framework



to each measure, whether independently or as parts of response scales, will be fully discussed in the results section, where we examine the extent to which TIMSS contextual questionnaires captured the learning-centered school condition framework. The items included and their coding can be found in Appendix Tables 1 and 2.

Noteworthy at this point is the fact that the TIMSS 2007 school and teacher questionnaires included a number of identical questions regarding school learning conditions. Thus, for the six main learning-centered conditions, we included these items to form four unique scales that allowed us to compare the perceptions of the principals and teachers in relation to

1. Teachers' expectations for student achievement;
2. Teachers' understanding of the curriculum goals and degrees of success in implementing the curriculum within the school;
3. Teachers' professional development in the school; and
4. Parents' and students' desire to do well in school.

We included control variables for background factors of schools, teachers, and students (Table 2). We carefully selected these variables after reference to the literature addressing the impact of available school resources, student SES, and home conditions (Baker et al., 2002; Heyneman & Loxley, 1983; Mullis et al., 2005). School characteristics refer to school-level background and composition factors that are typically considered as given conditions and therefore not malleable. In our study, these conditions included grade enrollment size, type of community (urban or rural), percentage of students from economically disadvantaged backgrounds, number of students in the class tested by TIMSS, and the percentage of students who took the mathematics assessment in their native language. Teacher characteristics included whether the teacher majored in mathematics, whether he or she was a certified teacher, gender, and years of teaching. Lastly, we included the student characteristics to account for the predictive association between SES and achievement: gender, language, home resources (calculator, computer, desk, dictionary, Internet, number of books), father's education, and mother's education.

We also controlled, to a limited extent, for instructional structure, including ability grouping, use of mathematics textbooks, minutes per week for mathematics teaching, and the amount of homework for mathematics per week. These instructional practice variables, which could be considered within the realm of influence of school management and leadership, are, we consider, more closely connected with classroom-level teacher practice and therefore outside the scope of this paper with its focus on school-level leadership and conditions.

Analytic Strategies

Our analytic strategies encompassed three steps. During Step 1, we tested the construction of key school-level variables based on the learning-centered leadership framework in an iterative process that accounted for both theoretical justification (validity) and internal reliability. We started by grouping all the TIMSS questionnaire items that might cover the construct domain of learning-centered school conditions based on the original construct maps for the TIMSS contextual questionnaires (Mullis et al., 2008). We then calculated the Cronbach's alpha to test the internal consistency of these constructs by system. In some cases where the clustering pattern had the potential to mask the nuanced differences among various learning-centered conditions, we calculated the internal consistency of each subscale separately. For example, the TIMSS 2007 school and teacher questionnaires contain eight items that measure "school climate for learning" (National Center for Education Statistics [NCES], 2008). We created three subscales to measure teachers' curriculum understanding and implementation (two items), teachers' expectations for achievement (one item), and culture of learning (four items). Items that did not cluster but fitted in the theoretical domains were also identified and grouped accordingly. We continued this process until we had examined all items of the two questionnaires for schools and teachers.

During Step 2, we aggregated teacher-level data to the school level. The ideal situation would be for teacher level to serve as a separate level from student level and school level, such that we could investigate independent teacher-level effects in a three-level hierarchical linear model (HLM). However, the small numbers of teachers (classrooms) per school in our dataset limited our options to conduct three-level HLM analyses. For example, Hong Kong SAR and Korea had only one mathematics teacher per school in the TIMSS 2007 data, Chinese Taipei had one mathematics teacher in 97.7% of its schools, and Japan had one mathematics teacher in 83.7% of its schools. Singapore had only one mathematics teacher in 0.7% of its schools, and the United States had one mathematics teacher in 6.3% of its schools. To maintain the comparability of our analysis results across the six selected education systems, we decided to aggregate

Table 2: Weighted means and standard deviations of control variables

Control variables	Chinese Taipei		Hong Kong SAR		Japan		Korea		Singapore		United States	
School level												
Grade 8 enrollment	588.08	(337.46)	192.58	(30.78)	157.57	(71.45)	368.9	(151.97)	312.73	(74.74)	253.43	(175.06)
Type of community	4.58	(1.10)	5.11	(0.84)	4.60	(1.16)	5.29	(0.94)	6.00	(0.00)	3.43	(1.53)
Economically disadvantaged	1.57	(0.81)	2.89	(1.07)	1.55	(0.72)	2.35	(1.03)	1.81	(0.95)	2.88	(1.06)
Tested in native language	2.80	(1.20)	3.82	(0.57)	4.00	(0.00)	4.00	(0.00)	1.45	(0.88)	3.43	(0.98)
Teacher qualifications												
Math major	0.85	(0.36)	0.79	(0.38)	0.86	(0.30)	0.95	(0.20)	0.79	(0.30)	0.69	(0.38)
Female	0.56	(0.50)	0.43	(0.47)	0.42	(0.45)	0.63	(0.43)	0.63	(0.38)	0.70	(0.38)
Teaching certificate	0.98	(0.14)	0.98	(0.14)	0.99	(0.08)	0.99	(0.08)	0.98	(0.12)	0.96	(0.15)
Years in teaching	12.12	(8.23)	13.18	(9.38)	15.46	(8.50)	13.57	(8.83)	8.34	(7.96)	13.98	(8.67)
Ability grouping	0.15	(0.36)	0.42	(0.50)	0.30	(0.46)	0.76	(0.43)	0.34	(0.47)	0.74	(0.44)
Classroom context												
Use math textbooks	0.93	(0.25)	1.00	(0.05)	0.98	(0.11)	0.97	(0.15)	0.90	(0.25)	0.92	(0.23)
Number of students in TIMSS class	34.79	(6.35)	36.31	(8.66)	33.29	(7.67)	36.7	(5.20)	37.28	(4.17)	22.35	(6.63)
Minutes/week for math teaching	238.32	(56.13)	247.27	(63.48)	157.43	(21.91)	181.34	(13.60)	218.56	(29.00)	247.77	(67.42)
Amount of homework per week	2.22	(0.83)	2.57	(0.70)	1.81	(0.82)	1.76	(0.75)	2.49	(0.62)	2.75	(0.58)
Length of homework per week	2.47	(0.73)	2.38	(0.52)	1.98	(0.74)	2.14	(0.64)	2.43	(0.65)	2.10	(0.47)
Student characteristics												
Female	0.47	(0.14)	0.50	(0.24)	0.50	(0.12)	0.48	(0.34)	0.49	(0.23)	0.50	(0.12)
Speaks language of test at home	2.33	(0.39)	2.63	(0.29)	2.93	(0.07)	2.68	(0.15)	1.62	(0.41)	2.67	(0.33)
Possesses calculator	0.98	(0.03)	0.99	(0.02)	0.98	(0.03)	0.96	(0.04)	0.99	(0.02)	0.96	(0.05)
Possesses computer	0.94	(0.07)	0.98	(0.03)	0.88	(0.09)	0.99	(0.02)	0.93	(0.07)	0.94	(0.07)
Possesses study desk	0.89	(0.09)	0.77	(0.12)	0.95	(0.06)	0.96	(0.04)	0.88	(0.09)	0.84	(0.09)
Possesses dictionary	0.98	(0.03)	0.98	(0.03)	0.99	(0.03)	0.99	(0.02)	0.98	(0.03)	0.91	(0.07)
Possesses Internet connection	0.89	(0.11)	0.97	(0.05)	0.77	(0.14)	0.96	(0.04)	0.86	(0.12)	0.86	(0.12)
Mother's education	2.24	(0.73)	1.90	(0.58)	2.97	(0.43)	2.90	(0.55)	2.37	(0.70)	3.38	(0.81)
Father's education	2.48	(0.82)	2.15	(0.69)	3.32	(0.57)	3.46	(0.62)	2.68	(0.78)	3.32	(0.84)
Number of books at home	2.92	(0.55)	2.49	(0.58)	2.97	(0.41)	3.46	(0.43)	2.81	(0.49)	2.95	(0.62)

teacher-level data at the school level. However, we needed to keep in mind the possibility of aggregation bias (Snijders & Bosker, 1999).

In situations where more than one teacher per school was surveyed, we calculated the school means of teacher-level variables. For the constructs, which were the same in the principal and teacher questionnaires, we calculated their correlations. The small correlations (usually $r < 0.30$) indicated that the principal- and teacher-reported constructs were different enough to be analyzed. This approach allowed us not only to focus on the effects of the school-level variables in our analysis but also to handle the complicated data structure arising out of more than one teacher teaching students in the same class in some schools.

During Step 3, we constructed the two-level hierarchical linear model (HLM) with students nested within schools to account for clustering data structure (Raudenbush & Bryk, 2002). This allowed us to examine the associations among the explanatory variables for learning-centered school conditions and student achievement, while controlling for covariates at the school, classroom, and student levels. We conducted the analysis by running the unconditional model that did not include any covariates and the conditional (full) model that included all explanatory and control variables formed from the student, teacher, and school questionnaires. We then calculated the intra-class correlations (ICCs) for the unconditional and conditional models, and the percentages of variance explained by the explanatory and control variables. We used the SAS PROC MIXED procedure to conduct the weighted HLM analysis.

Because we had five imputed datasets that corresponded to five plausible dependent variables, which we created using multiple imputation (Foy & Olson, 2009), we ran the models five times with five plausible variables, respectively, then used the SAS PROC MIANALYZE procedure to summarize the results. The detailed two-level HLM was as follows:

Level 1 (student): $y_{ij} = \beta_{0j} + \beta_{xj} X_{ij} + e_{ij} \quad e_{ij} \sim N(0, \sigma^2)$

Level 2 (school): $\beta_{0j} = \gamma_0 + \gamma_w W_j + u_j \quad u_j \sim N(0, \tau^2)$

$\beta_{xj} = \gamma_x$

The reduced model was:

$$y_{ij} = \gamma_0 + \gamma_w W_j + \gamma_x X_{ij} + u_j + e_{ij}$$

where y_{ij} is Grade 8 overall mathematics achievement in the IRT (item response theory) scale for student *i* in school *j*, and where W_j is a vector of school-level covariates for school *j*. X_{ij} is a vector of the Level-1 covariate for student *i* in school *j*. γ_w are the coefficients of school-level covariate, W_j, and γ_x are the coefficients of the Level-1 covariate, X_{ij}.

Our primary interests lay in interpreting the coefficients of the school-level explanatory variables. For the purpose of simplification, we used the fixed effects of student-level variables across schools. We then compared the means of explanatory variables and their coefficients, which represented the average effects of the explanatory variables on students' mathematics achievement across the selected systems. Given the cross-sectional nature of the TIMSS results and the contextual questionnaires, our methodological strategies addressed issues raised by Rutkowski et al. (2010) on causality claims, sampling, weights, proficiency estimates, imputed values, and generalization.

RESULTS

We first addressed this research question: How do the core components of learning-centered school conditions compare cross-culturally as described by the TIMSS contextual questionnaires? Our analyses of questionnaire content and construct factor patterns provided initial evidence that a set of variables reflecting the learning-centered leadership framework for school conditions might be identified with sufficient internal reliability using TIMSS teacher and school questionnaire items in the six selected systems. Some measures, however, appeared to have low internal reliabilities and some appeared to lack content validity due to limited coverage of the construct domain by the available questionnaire items.

Measures for learning-centered school conditions grouped into six sets:

1. Standards for learning;
2. Curriculum and instruction implementation;
3. Teacher professional community;
4. Parental involvement;
5. Performance accountability; and
6. Culture of learning.

Table 3 provides descriptive statistics for the key explanatory variables. There are several noteworthy findings. Under the main measure of c*urriculum and instruction implementation*, principals' direct efforts were measured by two items on the percentage of time spent on (a) "instructional leadership" and (b) teaching. The descriptive results show that the principal-reported time spent on activities related to instructional leadership varied across the six education systems. The United States average of 23.6% was lower than that of Korea (25.8%) and Chinese Taipei (25.0%), but was higher than that of Japan (22.9%) and Hong Kong SAR (19.9%).

The 2007 TIMSS questionnaire defines instructional leadership as "e.g., developing curriculum and pedagogy." The term may have been perceived very differently by the responding principals, however. In recent decades, the research-based focus on instructional leadership has led to the development of conceptual frameworks and instruments that aim to capture the complexity of this construct domain. For example, the widely used Principal Instructional Management Rating Scale (PIMRS; Hallinger, 1990, 2011; Hallinger & Murphy, 1985), which has 10 subscales and 50 items, proposes three dimensions to the role of an instructional leader—defining the school's mission, managing the instructional program, and promoting a positive school learning climate.

Table 3: Weighted means, standard deviations, and Cronbach's alphas for factors: explanatory variables from 2007 TIMSS contextual surveys

Explanatory variables	Chinese Taipei		Hong Kong SAR		Japan		Korea		Singapore		United States	
	Mean (*SD*)	Alpha	Mean (*SD*)	Alpha	Mean (*SD*)	Alpha	Mean (*SD*)	Alpha	Mean (*SD*)	Alpha	Mean (*SD*)	Alpha
Learning-centered leadership framework												
1. Standards for student learning												
Principal-reported expectation	4.11 (0.66)		3.80 (0.71)		3.53 (0.64)		3.81 (0.75)		3.84 (0.70)		3.97 (0.73)	
Teacher-reported expectation	3.91 (0.74)		3.58 (0.72)		3.45 (0.71)		3.69 (0.68)		3.62 (0.69)		3.95 (0.70)	
2. Curriculum and instruction												
Principal-reported curriculum rigor (2 items)	4.07 (0.47)	0.70	3.88 (0.46)	0.69	3.57 (0.57)	0.70	3.89 (0.53)	0.73	3.86 (0.49)	0.65	3.99 (0.63)	0.79
Teacher-reported curriculum rigor (2 items)	3.98 (0.55)	0.73	3.57 (0.56)	0.80	3.36 (0.54)	0.74	3.63 (0.49)	0.83	3.61 (0.48)	0.71	3.89 (0.59)	0.83
Time in instructional leadership (%)	25.00 (11.78)		19.91 (7.88)		22.86 (9.43)		25.83 (12.67)		21.35 (11.29)		23.56 (12.94)	
Time in teaching (%)	7.59 (10.1)		3.56 (7.84)		6.87 (8.24)		12.17 (12.67)		1.91 (2.92)		3.02 (6.62)	
3. Professional community												
Principal-reported PD (5 items)	2.99 (0.86)	0.87	3.55 (0.84)	0.86	2.84 (0.85)	0.77	2.77 (0.79)	0.86	4.25 (0.76)	0.86	4.09 (0.93)	0.89
Teacher-reported PD (6 items)	0.70 (0.29)	0.77	0.66 (0.31)	0.82	0.48 (0.27)	0.69	0.37 (0.28)	0.78	0.72 (0.22)	0.69	0.71 (0.27)	0.78
Teacher-reported collaboration (4 items)	1.67 (0.44)	0.64	1.81 (0.41)	0.60	1.95 (0.54)	0.60	1.73 (0.32)	0.54	1.83 (0.36)	0.63	1.77 (0.52)	0.73
4. External community												
Parental involvement (5 items)	0.77 (0.21)	0.48	0.78 (0.22)	0.46	0.59 (0.18)	0.16	0.62 (0.19)	0.26	0.83 (0.18)	0.32	0.92 (0.15)	0.37
5. Performance accountability												
Observation by principal	0.64 (0.48)		0.97 (0.17)		0.92 (0.27)		0.93 (0.26)		0.99 (0.08)		0.98 (0.14)	
Observation by external inspection	0.12 (0.32)		0.31 (0.47)		0.62 (0.49)		0.57 (0.50)		0.08 (0.28)		0.31 (0.46)	
Evaluated with student achievement	0.76 (0.43)		0.80 (0.40)		0.58 (0.50)		0.89 (0.32)		0.97 (0.16)		0.73 (0.45)	
Teacher peer review	0.33 (0.47)		0.70 (0.46)		0.50 (0.50)		0.83 (0.37)		0.53 (0.50)		0.27 (0.44)	
Incentive to recruit or retain teachers	0.06 (0.24)		0.05 (0.21)		0.18 (0.38)		0.10 (0.31)		0.14 (0.35)		0.06 (0.24)	
6. Culture of learning												
Principal's perception of parent and student desire to do well (4 items)	3.72 (0.61)	0.81	3.33 (0.59)	0.76	3.25 (0.56)	0.71	3.34 (0.63)	0.76	3.38 (0.57)	0.79	3.29 (0.72)	0.85
Teacher's perception of parent and student desire to do well (4 items)	3.11 (0.66)	0.84	2.98 (0.63)	0.81	3.02 (0.63)	0.71	2.88 (0.59)	0.78	2.93 (0.66)	0.85	2.93 (0.78)	0.90

On average, the Korean principals were those (from across the systems) who were most likely to be engaged in teaching (12.7%). The principals from Singapore appeared to be those least engaged in direct teaching (1.9%). The United States principals reported spending 3.2% of their time in direct teaching. Time spent in teaching could be affected by the size of school and the subject-matter training of the principal. Whether a principal carries a teaching workload may also depend on the career pathways of school personnel. For example, Singapore schools offer three separate tracks for career advancement, such that a teacher can aspire to be a master teacher, an administrator, or an instructional specialist with the Ministry of Education (Tucker, 2011).

The descriptive differences under *performance accountability* are interesting. When the principals were asked whether (yes or no) each of the four methods for evaluating teacher practice was being used in their schools, more than 95% of them in five of the systems (the exception was Chinese Taipei at 64%) said that they or senior staff used observations. However, external inspection practices appeared to be done very differently among the systems selected, with the range extending from rarely done in Singapore (8%) and Chinese Taipei (12%) through to being somewhat more regularly done in the United States (31%), and on to being considerably more common in Japan (62%) and Korea (57%). As many as 83% of the principals from Korea and 70% from Hong Kong SAR reported utilizing teacher peer review to evaluate mathematics practices, and more than 50% in Singapore and Japan reported the same. The United States principals reported the lowest occurrences (27%).

Table 3 also includes Cronbach's alphas for the variables constructed from the multiple questionnaire items. The alphas were calculated separately for each system. Overall, the constructed variables demonstrated high internal consistency, with the alphas ranging from 0.60 to 0.90. However, the internal consistency for the five items that formed the parental involvement scale appear to be in question across the six systems, given the consistently low Cronbach's alphas, which ranged from 0.16 in Japan to 0.48 in Chinese Taipei, and with the United States at 0.37. The five items (attend special event, raise funds, volunteer for school projects, ensure homework is completed, and serve on a school committee) may be covering very different construct dimensions of parental involvement and so do not tap into a common domain.

There were also discernible differences between the perceptions of the principals and the perceptions of the teachers across the six systems on the four sets of scales formed from the common items in the school and teacher questionnaires. With respect to teachers' expectations for student achievement in the school, principals reported higher average ratings (on a scale from 1 = very low to 5 = very high) than the teachers of the sampled classes in all five East Asian systems, while the average ratings were about the same in the United States. On teachers' understanding of curriculum goals and their degree of success in curriculum implementation, principals consistently reported higher ratings than the teachers for the sampled classes. The largest difference between principal and teacher perceptions was for the importance that learning held for parents and students. Here, teacher ratings (also on a scale from 1 = very low to 5 = very high) were significantly lower than the principal ratings across all six systems.

Table 4 shows the principal–teacher correlations for the four scale means (expectation, curriculum, professional development, and culture of learning). Among the six systems, teachers' and principals' viewpoints on learning culture appeared to be relatively highly correlated (correlations ranging from 0.27 in Korea to 0.60 in the United States). However, correlations on the other three measures were much lower. For example, correlations ranged from a low of 0.14 in Chinese Taipei to 0.38 in Singapore on teachers' expectations for student achievement, and from a negative 0.09 in Singapore to 0.28 in the United States. Given the low correlations, we decided to include the principal and teacher scale measures separately in the HLM analysis.

Table 4: Correlations of principals' and teachers' perceptions

Variable	Chinese Taipei	Hong Kong SAR	Japan	Korea	Singapore	United States
Expectation	0.14	0.30	0.26	0.17	0.38	0.25
Curriculum	0.07	0.20	0.30	0.13	0.24	0.35
Professional development	0.00	0.07	0.06	-0.03	-0.09	0.28
Culture of learning	0.36	0.36	0.42	0.27	0.53	0.60

We then addressed the second and third research questions regarding the associations between core components of learning-centered school conditions and student achievement in mathematics. The intra-class correlations (ICCs) and percentages of variance explained by the explanatory and control variables are reported in Table 5. The ICCs for the unconditional model varied from 0.09 to 0.64, with a mean of 0.33 across the six education systems, indicating considerable variation in the proportion of between-school mathematics achievement across the six systems; a high average of 33% of variation was due to between-school variation. The percentages of between-school variance explained by the explanatory and control variables varied from 54.5% to 74.6%, with a mean of 64.8% across the six systems, which meant they were relatively stable. These results suggest that these variables could account for more than 50% of between-school variation.

The fixed results of the full HLM (Table 6) show that some learning-centered school conditions were associated with student achievement in mathematics when background factors were held constant. However, which element of a learning-centered school condition and the extent to which the element was significant appeared to vary by national context. We will explain each condition specifically.

Table 5: Variances, ICC, and percentage of variance explained by the explanatory and control variables

Model	Variance and ICC	Chinese Taipei	Hong Kong SAR	Japan	Korea	Singapore	United States
Unconditional model[a]	Between school variance	2428.10	5892.37	1405.48	794.71	4108.09	2078.35
	Within school variance	8895.11	3351.38	5805.35	7690.23	4546.53	3787.35
	Total variance	11323.21	9243.75	7210.83	8484.94	8654.62	5865.70
	ICC	0.21	0.64	0.19	0.09	0.47	0.35
Conditional model[b]	Between school variance	807.52	2683.75	451.38	202.22	1451.65	816.61
	Within school variance	7646.14	3235.91	4923.08	6335.18	4116.13	3491.82
	Total variance	8453.66	5919.67	5374.46	6537.40	5567.78	4308.43
	ICC	0.10	0.45	0.08	0.03	0.26	0.19
Percentage variance explained[c]	Between school	66.7	54.5	67.9	74.6	64.7	60.7
	Within school	14.0	3.4	15.2	17.6	9.5	7.8
	Total variance	25.3	36.0	25.5	23.0	35.7	26.5

Notes:

a Unconditional model is the two-level HLM without any covariates.

b Conditional model is the two-level HLM including all variables in Tables 1 and 2.

c Percentage of variance explained was calculated by 100*(1-variance in conditional model/variance in unconditional model).

High Standards for Student Learning

Among the six selected systems, one point of increase in teachers' expectations on student achievement was positively associated with 7.77 points of increase ($p < 0.10$) in Japan and 22.63 points of increase ($p < 0.01$) in Singapore on the mathematics assessment scores. As Rutter and Jacobson (1986) have suggested, the perceptions that teachers have of student ability might affect their engagement in teaching and school improvement. Betts and Grogger (2003) found that, on average, higher grading standards were associated with higher Grade 12 test scores. However, among the selected systems, the positive associations were only statistically significant with respect to the teacher-reported measure, and for these two systems only.

Rigorous Curriculum and Instruction

We found that teachers' understanding of curriculum goals and their success in curriculum implementation at the school level did not appear to be strongly associated with the TIMSS 2007 Grade 8 students' mathematics achievement results. This finding held for both principal-reported and teacher-reported measures. In fact, there was a negative and statistically significant association between principal-reported curriculum-instruction implementation and student achievement results in Chinese Taipei (-12.97, $p < 0.10$). It will be interesting to further disentangle this association in terms of school, teacher, and student characteristics within the Chinese Taipei system.

Table 6: Learning-centered school conditions and mathematics achievement: fixed effects from two-level HLM

Explanatory variables	Chinese Taipei		Hong Kong SAR		Japan		Korea		Singapore		United States	
Learning-centered leadership framework												
1. Standards for student learning												
Principal-reported expectation	4.16	(5.93)	0.51	(10.43)	3.33	(4.7)	-4.00	(3.27)	2.46	(6.99)	4.65	(4.87)
Teacher-reported expectation	-2.90	(5.30)	5.91	(9.80)	7.77[+]	(4.42)	4.66	(3.44)	22.63[**]	(8.04)	-3.60	(4.66)
2. Curriculum and instruction												
Principal-reported curriculum rigor	-12.97[+]	(5.75)	22.15	(15.38)	5.08	(5.23)	4.47	(4.32)	-8.22	(9.85)	-2.88	(5.62)
Teacher-reported curriculum rigor	0.35	(6.78)	-1.44	(12.31)	-1.26	(5.05)	-6.17	(4.65)	9.31	(9.9)	-1.98	(5.08)
Time in instructional leadership (%)	0.24	(0.27)	-0.49	(0.88)	-0.54[*]	(0.26)	-0.01	(0.16)	-0.36	(0.32)	-0.25	(0.20)
Time in teaching (%)	0.15	(0.31)	0.13	(0.79)	-0.16	(0.3)	-0.05	(0.16)	0.35	(1.21)	-0.35	(0.38)
3. Professional community												
Principal-reported PD	-1.18	(4.05)	-2.37	(7.34)	0.56	(2.87)	1.35	(2.64)	3.53	(4.94)	0.19	(2.84)
Teacher-reported PD	-6.88	(11.47)	-10.43	(19.34)	-17.17[+]	(9.08)	0.04	(7.62)	-17.53	(17.4)	9.06	(10.28)
Teacher-reported collaboration	5.59	(7.77)	15.52	(15.45)	-11.63[*]	(4.98)	11.05	(6.88)	18.10[+]	(10.19)	0.59	(5.07)
4. External community												
Parental involvement	-1.65	(15.13)	-32.12	(29.13)	-4.96	(13.83)	-3.34	(10.57)	28.78	(21.47)	-3.84	(15.89)
5. Performance accountability												
Observation by principal	9.16	(7.88)	29.85	(36.22)	-28.45[**]	(9.64)	7.16	(8.08)	52.63	(50.93)	20.16	(17.87)
Observation by external inspection	5.51	(9.70)	-4.03	(13.27)	-4.29	(5.39)	-13.52[**]	(4.22)	2.62	(12.52)	-4.58	(5.42)
Evaluated with student achievement	4.82	(8.24)	3.18	(16.88)	-2.63	(4.87)	2.51	(6.48)	7.36	(24.84)	-5.96	(5.51)
Teacher peer review	7.76	(7.03)	4.14	(13.28)	-1.18	(4.75)	10.10[+]	(5.63)	-1.53	(7.52)	-3.95	(5.64)
Incentive to recruit or retain teachers	20.74	(14.1)	46.98	(32.58)	-6.24	(6.50)	0.83	(6.16)	8.22	(10.91)	-10.51	(10.5)
6. Culture of learning												
Principal's perception of parent and student desire to do well	4.09	(6.90)	10.75	(13.71)	9.93[+]	(5.56)	-0.97	(4.05)	26.98[**]	(10.00)	4.91	(5.54)
Teacher's perception of parent and student desire to do well	6.45	(6.49)	27.67[*]	(12.73)	3.77	(5.27)	2.42	(4.60)	5.23	(8.38)	12.54[*]	(5.36)

Table 6: Learning-centered school conditions and mathematics achievement: fixed effects from two-level HLM (contd.)

Control variables	Chinese Taipei		Hong Kong SAR		Japan		Korea		Singapore		United States	
School level												
Grade 8 enrollment	0.02^{+}	(0.01)	0.18	(0.22)	-0.01	(0.04)	-0.02	(0.02)	0.13*	(0.06)	0.04*	(0.02)
Type of community	2.05	(3.29)	5.52	(8.07)	0.04	(2.16)	1.20	(2.49)	na		-4.44*	(1.82)
Economically disadvantaged (%)	-2.38	(4.20)	-16.03*	(6.28)	-7.08^{+}	(3.84)	-4.87*	(2.43)	-1.28	(4.01)	-11.85**	(3.37)
Tested in native language (%)	-0.46	(2.82)	2.41	(10.59)	na		na		-3.36	(4.40)	2.03	(2.76)
Ability grouping	4.13	(9.44)	-29.10*	(12.58)	13.30*	(5.42)	-2.56	(5.04)	-1.78	(7.45)	8.70	(5.65)
Teacher qualifications												
Math major	4.61	(9.16)	-13.13	(15.85)	-1.82	(7.89)	-14.39	(10.34)	21.42^{+}	(11.80)	4.78	(6.81)
Female	4.05	(7.04)	7.98	(13.41)	9.81^{+}	(5.17)	4.06	(5.33)	8.31	(9.80)	-8.22	(6.41)
Teaching certificate	29.62	(23.52)	-28.00	(42.08)	na		na		7.16	(30.98)	-10.92	(16.96)
Years in teaching	0.11	(0.39)	0.06	(0.68)	-0.09	(0.28)	0.00	(0.24)	-0.27	(0.46)	-0.08	(0.30)
Classroom context												
Use math text books	0.40	(12.75)	na		-27.37	(21.84)	-5.33	(14.44)	6.33	(15.62)	-15.43	(11.33)
No. of students in TIMSS class	0.75	(0.58)	1.78*	(0.73)	0.62^{+}	(0.34)	0.31	(0.50)	-0.73	(1.05)	-0.34	(0.39)
Minutes/week for math teaching	0.09	(0.06)	-0.15^{+}	(0.09)	0.01	(0.11)	-0.20	(0.15)	-0.26*	(0.13)	0.05	(0.04)
Amount of homework per week	5.01	(3.84)	19.16*	(8.98)	-2.59	(3.06)	-3.92	(2.58)	11.99*	(6.06)	2.64	(4.81)
Length of homework per week	17.22**	(4.49)	19.27^{+}	(11.31)	4.47	(3.15)	2.28	(3.00)	11.08^{+}	(5.75)	13.46*	(5.60)

Table 6: Learning-centered school conditions and mathematics achievement: fixed effects from two-level HLM (contd.)

Control variables	Chinese Taipei		Hong Kong SAR		Japan		Korea		Singapore		United States	
Student characteristics												
Female	-3.05	(3.28)	-16.51**	(3.13)	-6.50*	(2.69)	-3.86	(2.95)	6.09*	(2.42)	-8.01**	(1.76)
Speaks language of test at home	17.84**	(2.38)	6.79**	(2.05)	24.01**	(3.76)	9.43**	(2.39)	-0.08	(1.34)	3.33+	(1.68)
Possesses calculator	26.42*	(12.00)	47.63**	(14.57)	16.79+	(9.97)	-1.67	(7.74)	43.62**	(11.27)	11.78**	(4.54)
Possesses computer	39.78**	(8.82)	-16.02+	(9.71)	5.01	(4.79)	17.53	(14.98)	10.55	(6.08)	7.27+	(4.37)
Possesses study desk	14.82**	(5.11)	-10.29**	(2.68)	16.22**	(5.61)	12.5	(8.15)	13.45**	(3.68)	4.10+	(2.39)
Possesses dictionary	42.16**	(11.90)	9.94	(10.31)	64.14**	(11.46)	82.13**	(12.27)	33.53**	(10.00)	3.71	(3.26)
Possesses Internet connection	4.44	(5.81)	16.84*	(6.98)	16.99**	(3.64)	58.38**	(8.51)	28.31**	(4.53)	1.00	(3.72)
Mother's education	-0.05	(1.52)	-0.13	(1.45)	3.08	(1.72)	2.90**	(1.11)	-0.87	(0.88)	1.13+	(0.58)
Father's education	6.46**	(1.44)	1.04	(1.46)	10.47**	(1.31)	5.42**	(1.16)	4.26**	(0.80)	3.12**	(0.78)
Number of books at home	17.21**	(1.38)	2.05+	(1.09)	9.35**	(1.10)	21.54**	(1.19)	7.90**	(1.03)	10.48**	(0.78)
Sample size												
School	143		106		141		144		155		192	
Student	3,830		3,040		4,151		4,072		4,351		5,859	

Notes:

Entries are coefficients (standard errors in parentheses).

na represents that either data were not available or the variable was the same within the system; + $p < 0.10$, *$p < 0.05$, **$p < 0.01$.

The extent to which principals spent their time on curriculum development and instructional quality (called "instructional leadership" in the TIMSS questionnaires) or directly engaged in teaching was only weakly associated with student achievement across all six selected systems. This finding is consistent with prior research that found little direct impact of principal instructional involvement on student achievement (Heck, 1993; Leithwood et al., 2004).

Teacher Professional Community

Both the principal-reported and teacher-reported measures for teacher professional development yielded statistically nonsignificant associations with student achievement, with the exception of Japan, where a one unit of increase of frequency in teacher-reported professional development was negatively associated with 17.17 ($p < 0.10$) of mathematics scores. It is conceivable that professional development may be used as additional training for teachers of low-performing students. Teacher collaboration, defined by discussions, joint lesson planning, and observations of peer teaching, was positively associated with student learning in Singapore (18.10, $p < 0.10$). However, the association was negative (-11.63, $p < 0.05$) in Japan. Again, it is possible that the same activities were being utilized for different purposes in the different settings.

Connections with External Community

The parental involvement measure did not appear to be associated with student achievement, perhaps because of the low internal reliability of the measure. In other words, the types of parental activities included in the questionnaires may be beneficial for a positive learning environment in various ways, but a composite score (average of the five dichotomous items) may not be directly associated with student achievement. Qualitative research may be necessary to fully account for the richness of community involvement as one of the key leadership effects on school improvement, as described by Chen (2008) and Pan and Yu (1999) in their qualitative studies of Taiwanese schools.

Performance Accountability

The performance evaluation measures for teacher practices yielded largely insignificant coefficients on student achievement across the six systems, except for Japan and Korea. In Japan, being observed by the principal or senior staff for mathematics teaching evaluation was negatively associated with student achievement by 28.45 ($p < 0.01$) points of decrease. In Korea, being observed by external inspectors was also negatively associated with achievement (-13.52, $p < 0.01$). However, teacher peer review as a form of evaluation had a positive association with student achievement (10.10, $p < 0.10$) in Korea. It may be that external observation, whether by the principal, senior staff, or inspectors, reinforces standards and enhances instructional quality, especially for teachers with low-performing students or who need to engage in relevant professional development. On the other hand, peer review might occur when a teacher is considered exemplary. It would be interesting to examine if and the extent to which evaluation patterns are associated with teacher qualifications in our next steps of data analysis.

Culture of Learning

The principal-reported and teacher-reported measures that capture students' (and their parents') desire to do well produced strong and large coefficients on student achievement in all six systems, but especially in the United States (12.54, $p < 0.05$), Hong Kong SAR (27.67, $p < 0.05$), and Singapore (26.98, $p < 0.01$). While the finding here is consistent with the theoretical assumptions about cultural values relating to education and student achievement (Paik, 2004; Shen, 2005; Wang, 2004), the four items that formed the measure are broadly defined: parental support for student achievement, parental involvement in school activities, students' regard for school property, and students' desire to do well in school. These dimensions may be confounded with other explanatory factors. Culture of learning as a construct domain will need to be further extricated in order to identify elements that are malleable by principals, teachers, and other key stakeholders.

Control variables at the school, classroom, and student levels yielded results largely consistent with previous literature and theoretical assumptions. For example, the percentage of economically disadvantaged students in the school, being female, not tested in the native language, and lack of Internet access at home were negatively associated with student achievement. Moreover, frequency of mathematics assignments and length of time for homework were positively associated with mathematics scores when all other factors were held constant. The association between ability grouping and mathematics achievement varied among the selected systems; it was a positive one in Japan (13.30, $p < 0.05$), but negative in Hong Kong SAR (-29.10, $p < 0.05$). It is plausible that ability grouping as a practice is used for different purposes, ranging from addressing the diverse needs of low-performing students (and thus being a negative association with student achievement) to meeting the needs of advanced students (and thus a positive association).

DISCUSSION

The TIMSS contextual questionnaires have been recognized for providing important background information on the learning conditions of students in participating education systems. In this study, we sought a fuller appreciation of how TIMSS data might be used to inform the field of school administration and leadership.

Our findings indicate that the TIMSS contextual questionnaires provide an interesting crossnational snapshot of learning conditions in the participating nations. We found this snapshot multidimensional and informative. Specifically, by using the modified learning-centered leadership framework to identify items from the questionnaires that reflected the core components of malleable learning conditions, we were able to tap into all six dimensions of the main construct of learning-centered leadership, albeit not fully on all fronts. We then explored the cross-national evidence for the associations between core learning conditions within the realm of influence of school leadership and student achievement by connecting mathematics achievement outcomes with contextual questionnaire results. We also controlled for factors that might confound the relationships between school learning conditions and educational outcomes at multiple contextual levels.

The results showed that a number of learning conditions pertaining to accountability measures, classroom instructional practices, and attitudes toward learning were significantly and strongly associated with student learning in many of the selected systems. Our analyses of the selected systems also affirmed the notion that school leadership cannot simply be measured by the amount of direct instruction-related efforts of the principals (Hallinger & Heck, 2010; Leithwood & Jantzi, 1999; Louis et al., 2010). In essence, we operationalized, to an extent, the theoretical assumptions regarding leadership, learning-centered school conditions, and student achievement results within the TIMSS data framework. However, our probe also underscored the challenge of using TIMSS questionnaires to study the role of school leadership in student learning.

First, the TIMSS questionnaires did not appear to cover the full domain of the learning-centered leadership framework. The items available from the school, teacher, and student questionnaires were designed with minimizing the survey burden to respondents in mind. For example, only two items are available to reflect *standards for student learning* at the school level, and the items for connections with *external community* reflect parental involvement only.

Second, the items tend to be written in generic terms and lack the necessary specificity to describe what leaders should do to impact school conditions. For example, instructional leadership is defined as "developing curriculum and pedagogy," which is a very obvious simplification of the complex actions involved in providing this type of leadership (Hallinger & Murphy, 1987). Such vagueness may present a threat to the construct validity of the measures.

Time spent on classroom teaching by the principal is another example that deserves further consideration. Teaching may well afford principals the opportunity to be directly involved in classroom interaction with students. However, if schools have head teachers who take the lead in lesson planning and pedagogy development, principals may then be able to have a larger impact because of having the time to exercise instructional leadership at the school level OECD, 2010; Tucker, 2011).

In fact, the definitions and applications of learning-centered leadership will most certainly vary from country to country. Given that TIMSS 2007 questionnaires offer limited content validity on important learning conditions, in-depth and qualitative probes into how the same term is applied to these diverse definitions and practices may be necessary. For example, more in-depth understanding of what learning culture as a learning condition entails is needed beyond the items that we could identify from TIMSS in this study.

Third, the cross-sectional nature of the questionnaires and mathematics achievement provides limited insight into the relationship between the actions of school personnel in ensuring optimal learning conditions and student learning outcomes. When crossnational analysis shows opposing signs or varying degrees of associations, little explanation can be offered without more grounded investigation into the contexts of the schools and classrooms. Furthermore, even though the five East Asian systems selected for this study had average scores in mathematics that were statistically significantly higher than the corresponding scores in the United States, small point differences do not necessarily say something about the quality differences in education. If the influence of learning-centered leadership is to be analyzed in general—or in comparison with the United States—then the low-achieving systems should be chosen for reasons of validity, in the future. Overall, the limitations of this study underscore the need not only to broaden but to deepen our understanding of variation in educational and social contexts across countries so that we can fully appreciate the utility of international benchmarking for student achievement.

The 2008 NGA and CCSSO reports called for international benchmarking and proposed revising "state policies for recruiting, preparing, developing, and supporting teachers and school leaders to reflect the human capital practices of top performing nations and states around the world" (NGA, 2008, p. 27). The challenge, however, is to identify effective practices that take multilevel contextual factors into consideration. Using the TIMSS crossnational datasets from 2007, our study set out to examine the viability of using international assessment reports to inform school leadership practices. While our analysis did not include all participating nations, the results for the selected systems reveal interesting differences in school-level conditions for learning and how such conditions are associated with mathematics achievement. We hope to deepen our probe into available empirical evidences and identify convergent and divergent themes as compared with theoretical assumptions in the field.

Future research could further investigate the nature of linkage between school conditions malleable by leadership by:

(a) Including more countries in the analysis in order to identify any systematic differences between high- and low-achieving nations in the learning-centered conditions;

(b) Selecting a few countries of interest and conducting cohort comparisons among the four sets of TIMSS results (1995, 1999, 2003, and 2007) to determine if there are changes regarding important core learning-centered conditions and their associations with student learning over time;

(c) Analyzing local policies relevant to the design and implementation of improving learning-centered conditions; and

(d) Conducting analyses on whether learning-centered conditions are different with regard to the disaggregation of student subpopulations and school types, such as performance quartiles, racial and ethnic groups, public schools serving students with different socioeconomic concentrations, and urban versus rural schools.

To the extent that such insight identifies leadership components that are within the realm of control of school principals, information about the underlying processes gained from this study could be useful in informing cross-national research on school leadership regarding the development or the modification of existing professional training and evaluation. For the United States, such an approach might lead to the high levels of student achievement that other countries currently experience.

Appendix Table 1: Measures for learning-centered school conditions from 2007 TIMSS teacher questionnaire (TQ) and principal questionnaire (PQ)

Measures	Questionnaire items from TIMSS
	Standards for student learning
Teachers' academic expectation (TQ and PQ)	How would you characterize teachers' expectations for student achievement?
	Curriculum and instruction implementation
Curriculum rigor (TQ and PQ, 2 Items)	How would you characterize: • Teachers' understanding of the school's curricular goals within your school? • Teachers' degree of success in implementing the school's curriculum within your school?
Percentage of time in instructional leadership (PQ) Percentage of time in teaching (PQ)	By the end of this school year, approximately what percentage of time in your role as principal will you have spent on: • Instructional leadership (e.g., developing curriculum and pedagogy)? • Teaching?
	Teacher professional community
Professional development (PQ, 5 Items)	During this school year, how often have your eighth-grade teachers been involved in professional development opportunities for mathematics and science: • For mathematics and science targeted at supporting the implementation of the national or regional curriculum? • Targeted at designing or supporting the school's own improvement goals? • Targeted at improving content knowledge? • Targeted at improving teaching skills? • Targeted at using information and communication technology for educational purposes?
Professional development (TQ, 6 Items)	In the past two years, have you participated in professional development in: • Mathematics content? • Mathematics pedagogy/instruction? • Mathematics curriculum? • Integrating information technology into mathematics? • Improving students' critical thinking or problem-solving skills? • Mathematics assessment?
Teacher collaboration (TQ, 4 Items)	How often do you: • Have discussions about how to teach a particular concept with other teachers? • Have worked on preparing instructional materials with other teachers? • Visited another teacher's classroom to observe his/her teaching? • Have informal observations of your classroom by another teacher?

Appendix Table 1: Measures for learning-centered school conditions from 2007 TIMSS teacher questionnaire (TQ) and principal questionnaire (PQ) (contd.)

	External Community
Parental involvement (PQ, 4 Items)	Does your school expect parents to: • Attend special events (e.g., science fair, concert, sporting events)? • Raise funds for the school? • Volunteer for school projects, programs, and trips? • Ensure that their child completes his/her homework? • Serve on school committees (e.g., select school personnel, review school finances)?
	Performance accountability
Observation by principal	Are observations by the principal or senior staff used to evaluate the practice of eighth-grade mathematics teachers?
Observation by external inspection	Are observations by inspectors or other persons external to the school used to evaluate the practice of eighth-grade mathematics teachers?
Evaluated with student achievement	Is student achievement used to evaluate the practice of eighth-grade mathematics teachers?
Teacher peer review	Is teacher peer review used to evaluate the practice of eighth-grade mathematics teachers?
Incentive to recruit or retain teachers	Does your school currently use any incentives to recruit or retain eighth grade teachers in mathematics?
	Culture of learning
Parent and student desire to do well (PQ and TQ, 4 items)	How would you characterize Parental support for student achievement within your school?
	Parental involvement in school activities within your school?
	Students' regard for school property within your school?
	Students' desire to do well in school within your school?

Note: *Internal consistency, measured by Cronbach's alpha, for each scale specific to each system can be found in Table 2.

Appendix Table 2: Coding of measures

Variables	Coding
Learning-centered leadership framework	
1. Standards for student learning	
Principal-reported expectation	1= very low, 2 = low, 3 = medium, 4 = high, 5 = very high
Teacher-reported expectation	1= very low, 2 = low, 3 = medium, 4 = high, 5 = very high
2. Curriculum and instruction	
Principal-reported curriculum rigor	1= very low, 2 = low, 3 = medium, 4 = high, 5 = very high
Teacher-reported curriculum rigor	1= very low, 2 = low, 3 = medium, 4 = high, 5 = very high
Percentage time leadership	Percentage of time
Percentage time in teaching	Percentage of time
3. Professional community	
Principal-reported PD	1= none, 2 = 1–25%, 3 = 26–50%, 4 = 51–75%, 6 = 76–100 %
Teacher-reported PD	1 = yes, 0 = no
Teacher-reported collaboration	1= never or almost never, 2 = 2 or 3 times per month, 3 = 1–3 times per week, 4 = daily or almost daily
4. External community	
Parental involvement	1 = yes, 0 = no
5. Performance accountability	
Observation by principal	1 = yes, 0 = no
Observation with external inspection	1 = yes, 0 = no
Evaluated with student achievement	1 = yes, 0 = no
Teacher peer review	1 = yes, 0 = no
Incentive to recruit or retain teachers	1 = yes, 0 = no
6. Culture of learning	
Principal's perception of parent and student desire to do well	1= very low, 2 = low, 3 = medium, 4 = high, 5 = very high
Teacher's perception of parent and student desire to do well	1= very low, 2 = low, 3 = medium, 4 = high, 5 = very high

Appendix Table 2: Coding of measures (contd.)

Variables	Coding
Control variables	
School level	
Grade 8 enrollment	Number of students
Type of community	1 = 3,000 people or fewer, 2 = 3,001 to 15,000 people, 3 = 15,001 to 50,000 people, 5 = 100,001 to 500,000 people, 6 = more than 500,000 people
Percentage economically disadvantaged	1 = 0 to 10%, 2 = 11–25%; 3 = 26–50%, 4 = more than 50%
Percentage tested in native language	1 = less than 50%, 2 = 51–75%, 3 = 26–50%, 4 = more than 50%
Ability grouping	1 = yes, 0 = no
Teacher qualifications	
Math major	1 = yes, 0 = no
Female	1 = female, 0 = male
Teaching certificate	1 = yes, 0 = no
Years in teaching	Years
Classroom context	
Use math textbooks	1 = yes, 0 = no
Number of students in TIMSS class	Number of students
Minutes per week for math teaching	Minutes
Amount homework per week	0 = some homework, 1 = some lessons, 2 = about half the lessons, 3 = every or almost every lesson
Length of homework per week	1 = fewer than 15 minutes, 2 = 15–30 minutes, 3 = 31–60 minutes, 4 = 61–90 minutes, 5 = more than 91 minutes
Student characteristics	
Female	1 = female, 0 = male
Speak language of test at home	0 = never, 1 = sometimes, 2 = almost always, 3 = always
Possesses calculator	1 = yes, 0 = no
Possesses computer	1 = yes, 0 = no
Possesses study desk	1 = yes, 0 = no
Possesses dictionary	1 = yes, 0 = no
Possesses Internet connection	1 = yes, 0 = no
Mother's education	0 = ISCED Level 1 or 2, or did not go to school, 1 = ISCED 2; 2 = ISCED 3, 3 = ISCED 4, 4 = ISCED 5B, 5 = ISCED 5A, first degree, 6 = beyond ISCED 5A, first degree
Father's education	0 = ISCED Level 1 or 2, or did not go to school, 1 = ISCED 2; 2 = ISCED 3, 3 = ISCED 4, 4 = ISCED 5B, 5 = ISCED 5A, first degree, 6 = beyond ISCED 5A, first degree
Number of books at home	1 = 0–10, 2 = 11–25, 3 = 26–100, 4 = 101–200, 5 = over 100 books

References

Akiba, M., LeTendre, G. K., & Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, *36*(7), 369–387.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley effect" on mathematics and science achievement. *Comparative Education Review*, *46*(3), 291–312.

Baker, D. P., Lee, J., & Heyneman, S. P. (2003). Should America be more like them? Cross-national high school achievement and U.S. policy. *Brookings Papers on Education Policy*, *6*, 309–338).

Betts, J. R., & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, *22*, 343–352.

Chen, P. (2008). Strategic leadership and school reform in Chinese Taipei. *School Effectiveness and School Improvement*, *19*(3), 293–318.

Chudgar, A., & Luschei, T. F. (2009). National income, income inequality, and the importance of schools: A hierarchical cross-national comparison. *American Educational Research Journal*, *46*(3), 626–658.

Clarke, D., Mesiti, C., O'Keefe, C., Xu, L. H., Jablonka, E., Mok, I. A. C. ... Shimuzu, Y. (2007). Addressing the challenge of legitimate international comparisons of classroom practice. *International Journal of Educational Research*, *46*(5), 280–293.

Cobb, P., & Smith, T. (2008). District development as a means of improving mathematics teaching and learning at scale. In K. Krainer & T. Wood (Eds.), *International handbook of mathematics teacher education: Vol. 3. Participants in mathematics teacher education: Individuals, teams, communities and networks* (pp. 231–254). Rotterdam, the Netherlands: Sense Publishers.

Cohen, D., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, *102*, 294–343.

Council of Chief State School Officers (CCSSO). (1996). *Interstate school leaders licensure consortium standards for school leaders*. Washington, DC: Author.

Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instructions. *School Effectiveness and School Improvement*, *7*, 197–228.

Desimone, L. M. (2006). Consider the sources: Response differences among teachers, principals and districts on survey questions about their education policy environment. *Educational Policy*, *20*(4), 640–676.

Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.

Ferraro, D., & Van de Kerckhove, W. (2006). *Trends in International Mathematics and Science Study (TIMSS) 2003: Nonresponse bias analysis*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Foy, P., & Olson, J. F. (2009). *TIMSS 2007 user guide for the international database*. Boston, MA: Boston College.

Givvin, K. B., Hiebert, J., Jacobs, J. K., Hollingsworth, H., & Gallimore, R. (2005). Are there national patterns of teaching? Evidence from the TIMSS 1999 Video Study. *Comparative Education Review*, *49*(3), 311–343.

Glasman, N. S., & Heck, R. H. (1992). The changing leadership role of the principal: Implications for principal assessment. *Peabody Journal of Education*, *68*(1), 5–24.

Goldring, E., & Cravens, X. C. (2007). Teachers' academic focus on learning in charter and non-charter schools. In M. Berends, M. G. Springer, & H. J. Walberg (Eds.), *Charter school outcomes*. New York, NY: Lawrence Erlbaum Associates.

Goldring, E., Porter, A., Murphy, J., Elliott, S. N., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools*, *8*, 1–36.

Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., & Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Hallinger, P. (1990). *Principal Instructional Management Rating Scale*. Sarasota, FL: Leading Development Associates.

Hallinger, P. (2011). A review of three decades of doctoral studies using the Principal Instructional Management Rating Scale: A lens on methodological progress in educational leadership. *Educational Administration Quarterly*, *47*(2), 271–306.

Hallinger, P., & Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly*, *32*(1), 5–44.

Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership & Management*, *30*(2), 95–110.

Hallinger, P., & Murphy, J. (1985). Assessing the instructional leadership behavior of principals. *Elementary School Journal*, *86*, 217–248.

Hallinger, P., & Murphy, J. (1987). Instructional leadership in the school context. In W. D. Greenfield (Ed.), *Instructional leadership: Concepts, issues, and controversies* (pp. 179–203). Boston, MA: Allyn & Bacon.

Heck, R. H. (1993). School context, principal leadership, and achievement. *The Urban Review*, *25*(2), 151–166.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary school quality on academic achievement across 29 high- and low-income countries. *American Journal of Sociology*, *88*(6), 1162–1194.

Leithwood, K., & Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly*, *35* (Supplemental), 679–706.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning: A review of research for the Learning from Leadership Project*. New York City, NY: Wallace Foundation

LeTendre, G., Baker, D., Akiba, M., Goesling, B., & Wiseman, A. (2001). Teachers' work: Institutional isomorphism and cultural variation in the U.S., Germany, and Japan. *Educational Researcher*, *30*(6), 3–15.

LeTendre, G., Baker, D. P., Wiseman, A., Boe, E., & Goesling, B. (2002). *Classroom implementation of national curricula and cross-national patterns of achievement* (Vol. 19, Working Paper Series, Pennsylvania State University, Education Policy Studies). University Park, PA: Pennsylvania State University.

Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Learning from leadership: Investigating the links to improved student learning*. New York, NY: Wallace Foundation.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grade*s. Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., & Gonzalez, E. J. (2005). *TIMSS assessment frameworks and specifications 2003*. College Hill, MA: Boston College.

Murphy, J., Goldring, E., Elliott, S. N., & Porter, A. (2006). *Learning-centered leadership: A conceptual foundation*. New York, NY: Wallace Foundation

National Center for Education Statistics (NCES). (2008). *TIMSS 2007 assessment frameworks*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Governors Association (NGA). (2008). *Benchmarking for success: Ensuring U.S. students receive a world-class education*. Washington, DC: Author.

Organisation for Economic Co-operation and Development (OECD). (2010). *Strong performers and successful reformers in education: Lessons from PISA for the United States*. Paris, France: Author. Retrieved from http://www.pisa.oecd.org/dataoecd/32/50/46623978.pdf

Paik, S. J. (2004). Korean and U.S. families, schools, and learning. *International Journal of Educational Research*, *41*(1), 71–90.

Pan, H. L., & Yu, C. (1999). Educational reforms with their impacts on school effectiveness and school improvement in Chinese Taipei, R.O.C. *School Effectiveness and School Reform*, *10*(1), 72–85.

Porter, A., & Gamoran, A. (2002). *Methodological advances in cross-national surveys of educational achievement*. Washington, DC: National Academy Press.

Porter, A., Goldring, E., Murphy, J., Elliott, S. N., & Cravens, X. (2006). *A conceptual framework for the assessment of principal and team school leadership*. New York, NY: Wallace Foundation.

Provasnik, S., Gonzales, P., & Miller, D. (2009). *U.S. performance across international assessments of student achievement: Special supplement to "The Condition of Education 2009"* (NCES 2009-083). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Rice, R. C., & Islas, M. R. (2001). TIMSS and the influence of instructional leadership on mathematics and science performance. *NASSP, 85*(623), 5–9.

Rowan, B., Correnti, R., Miller, R. J., & Camburn, E. M. (2009). *School improvement by design: Lessons from a study of comprehensive school programs*. Madison, WI: Consortium for Policy Research in Education.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, *39*(2), 142–151.

Rutter, R., & Jacobson, J. (1986). *Facilitating teacher engagement*. Madison, WI: National Center on Effective Secondary Schools.

Schmidt, W. H., Rotberg, I. C., & Siegel, A. (2003). Too little too late: American high schools in an international context. *Brookings Papers on Education Policy*, *6*, 253–307.

Shen, C. (2005). How American middle schools differ from schools of five Asian countries: Based on cross-national data from TIMSS 1999. *Educational Research and Evaluation: An International Journal on Theory and Practice*, *11*(2), 179–199.

Smith, T., Desimone, L., & Ueno, K. (2005). "Highly qualified" to do what? The relationship between NCLB teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. *Educational Evaluation and Policy Analysis*, *27*(1), 75–109.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Swanson, C. B., & Barlage, J. (2006). *Influence: A study of the factors shaping education policy*. Bethesda, MD: Editorial Projects in Education.

Tucker, M. (2011). *Standing on the shoulders of giants: An American agenda for education reform*. Washington, DC: National Center on Education and the Economy.

U.S. Department of Education. (2010). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. Washington, DC: Office of Planning, Evaluation, and Policy Development.

Wang, D. B. (2004). Family background factors and mathematics success: A comparison of Chinese and U.S. students. *International Journal of Educational Research*, *41*(1), 40–54.

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*(5), 3–13.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells the U.S. about the effects of leadership on student achievement*. Aurora, CO: Mid-continent Research for Education and Learning.