The Achievement Gap, Revisited: An Empirical Assessment of What We Can Learn from East Asian Education

Professor Mary C. Brinton, Advisor Katherine J.D. Czehut

The Achievement Gap, Revisited:
An empirical assessment of what we can learn from East Asian education

Abstract

International mathematics assessments have established students in East Asia as among the best in the world and their U.S. counterparts as mediocre. What is not clear is why this "achievement gap" exists. The last major study to address this question, Stevenson and Stigler's (1992) *The Learning Gap*, was published prior to empirical and methodological advances in international comparative research on education. Prevailing wisdom points to unverified differences in cultural beliefs, which often leads to defeatist conclusions. This dissertation offers a fresh perspective by applying sociological theory and methods to the issue. Specifically, I rely on underutilized data from the 2003 and 2007 Trends in International Mathematics and Science Study (TIMSS) of fourth graders to compare educational systems across three major factors that influence math achievement: curriculum, teachers and parents.

My main empirical findings are that there is greater uniformity of math instruction across classrooms in the participating East Asian nations of Hong Kong, Japan, Singapore and Taiwan than in the U.S. and that, among all participating educational systems, average achievement tends to be higher in those with greater uniformity of instruction. The implication is that the institutional arrangements that allow for less uniformity of instruction across classrooms in the U.S. might be partially responsible for the gap. Cross-regional differences in teacher effectiveness might also account for part of the gap, as three-level, hierarchical linear models of achievement in each nation indicate that U.S. math teachers are less effective than their East Asian counterparts—even after the quantity of instruction provided is taken into account. The main theoretical contribution is an alternative explanation for the apparent cross-regional

disparity in the proportion of involved parents, which highlights how schools can make a difference in whether or not parents become involved. Such an approach promises a way out of the dead-end reached by previous theorists.

However, this dissertation also draws attention to the limitations of the existing data. At present, there is not enough information available to substantiate the policy recommendations made in previous studies. As such, a central aim of this dissertation is to put research onto sounder methodological footing.

Table of Contents

Acknowledgements

One of the first sayings I learned in Japanese was "okage sama desu," which is used to acknowledge that one's good fortune and wellbeing depends on the generosity and kindness of others. It is thanks to the following individuals that I have the great fortune of completing this dissertation: my advisor and mentor, Mary C. Brinton, whose patience and guidance got me through graduate school and whose wisdom and kindness transformed me into a scholar; Peter V. Marsden, who provided me with excellent and insightful comments and championed my research from the start; Jason Beckfield, whose provocative questions helped to shape the tenor of this dissertation; Christopher Winship, who took the time to answer my questions long after any formal obligations would have required a response; the Benjamin Bainbridge Tregoe fund, which generously sponsored the final year of my dissertation research; the Asia Center at Harvard University, which helped to fund a fact-finding trip to Singapore and Hong Kong for this research; the members of my cohort— Jeff Denis, Crystal Fleming, Ethan Fosse, Patrick Hamm, Onoso Imoagene, Ji Wook Jung, Matthew Kaliner, Una Kim, Charles Loeffler, Sarah Halpern Meekin, Eun Mi Mun, Laura Tach, Van Tran, Jessica Welburn, Shawna Bowden Vican and Min Zhou—who challenged me to rise to the occasion and commiserated with me during our collective quest for scholarly advancement; Shawna Vican, who has been by my side for the highs and the lows and without whom I never would have met my husband, David, who is now my biggest cheerleader and supporter; my parents, Hal and Kathy Drake, and especially my father, who spent countless hours on the phone helping me get my ideas across. To all of you: *okage sama desu*.

List of Tables

List of Figures

Chapter 1

Introduction

In December 2010, the 2009 round of the Program for International Student Assessment, or PISA, became the latest international assessment in mathematics to publish its results. PISA 2009 is the 10th international assessment in math that the U.S. has participated in since the 1960s, and because most studies assess student achievement at several different grade levels, it represents the 19th opportunity the U.S. has had to crack the list of top five nations in math. It also represents the 19th time that the U.S. has failed to do so, even though President George H. W. Bush made it a national priority to reach the No. 1 spot in both math and science back in 1990 (Vinovskis 1999). The average math score of 15-year-old U.S. students on the 2009 PISA is below the average for the 32 OECD member nations that participated and ranks 31st out of 65 educational systems. Facing reporters regarding the 2009 PISA results, Secretary of Education Arne Duncan was blunt in his response: “We have to see this as a wake-up call,” he said, adding: “We can quibble, or we can face the brutal truth that we’re being out-educated.”[1]

International assessments such as PISA and the Trends in International Mathematics and Science Study, or TIMSS, are designed to tell us how educational systems perform compared to

[1] Quote excerpted from Dillon (2010).

one another. At this point, the pattern of mediocrity in U.S. education is well-established. What is also clear is that East Asian nations are the best in the world at providing their youth with a superior math education. Since the 1980s, East Asian nations consistently have topped the list of high-achieving nations in math, as the summary of international math achievement scores from the U.S. and East Asian nations provided in Table 1.1 indicates. This is no small feat, especially considering that the number of nations participating in each of these studies has ballooned from 18 to almost 70. These results suggest that East Asian educational systems—in particular those of Hong Kong, Japan, Singapore, South Korea and Taiwan—are more effective than ours at imparting math knowledge.

Although the existence of this "achievement gap" is obvious, the reasons for it are far less clear. Scholars and laypeople alike tend to focus on East Asian culture as the explanation, and in the process rarely consider what educational systems in East Asian nations are doing to promote greater achievement. In this dissertation, I will argue that the role of educational inputs is far more important than the role that cultural legacies might play. The difference in approach is significant: to change our national culture would be a task far beyond the capacity of any department of education, teacher-training program or textbook publisher, whereas modifying instructional materials, teacher-certification requirements or school calendars is a reasonably attainable goal.

Table 1.1 Summary of U.S. and East Asian Performance on International Assessments in Math

	Year	Survey Name	Target Population	N Edu. Systems	Participating East Asian Nations	Rank of the U.S.	Rank of East Asian Nations
1	1964	First International Mathematics Study (FIMS)[a]	13-year-olds	12	Japan	10-11	Japan: 1-2
2	1964	FIMS	Students in final year of secondary school	12	Japan	12	Japan: 5-6
3	1980-2	Second International Mathematics Study (SIMS)	13-year-olds	20	Hong Kong and Japan	Arithmetic- 10, Algebra- 8-14, Geometry- 13-17, Measurement- 18, Descriptive Statistics- 7-13	Arithmetic- Japan: 1-4, Hong Kong: 3-7, Algebra- Japan: 1, Hong Kong: 8-13, Geometry- Japan: 1, Hong Kong: 7-11, Measurement- Japan: 1, Hong Kong: 7-10, Descriptive Statistics- Japan: 1, Hong Kong: 7-14
4	1980-2	SIMS	Students in final year of secondary school	15	Hong Kong and Japan	Numbers- 11-13, Algebra- 13-14, Geometry- 12-14, Functions and Calculus- 12	Numbers- Hong Kong: 1, Japan: 2 Algebra- Hong Kong: 1-2, Japan: 1-2, Geometry- Hong Kong: 1, Japan: 2, Functions and Calculus- Hong Kong: 1, Japan: 2
5	1988	First International Assessment of Educational Progress (FIAEP)	13-year-olds	12	Korea	12	Korea: 1
6	1995	Third International Mathematics and Science Study (TIMSS)	Third graders	24	Hong Kong, Japan, Korea and Singapore	10	Korea: 1, Singapore: 2, Japan: 3, Hong Kong: 4

Table 1.1 Continued

	Year	Survey Name	Target Population	N Edu. Systems	Participating East Asian Nations	Rank of the U.S.	Rank of East Asian Nations
7	1995	TIMSS	Fourth graders	26	Hong Kong, Japan, Korea and Singapore	12	Singapore: 1, Korea: 2, Japan: 3, Hong Kong: 4
8	1995	TIMSS	Seventh graders	39	Hong Kong, Japan, Korea and Singapore	24	Singapore: 1, Korea: 2, Japan: 3, Hong Kong: 4
9	1995	TIMSS	Eighth graders	41	Hong Kong, Japan, Korea and Singapore	28	Singapore: 1, Korea: 2, Japan: 3, Hong Kong: 4
10	1995	TIMSS	Students in final year of secondary school	21	None	19	
11	1999	TIMSS	Eighth graders	38	Hong Kong, Japan, Korea, Singapore and Taiwan	19	Singapore: 1, Korea: 2, Taiwan: 3, Hong Kong: 4, Japan: 5
12	2000	Program for International Student Assessment (PISA) and PISA+[b]	15-year-olds	41	Hong Kong, Japan and Korea	20	Hong Kong: 1, Japan: 2, Korea: 3
13	2003	PISA[c]	15-year-olds	41	Hong Kong, Japan, Korea and Macau-China	25-28	Hong Kong: 1-3, Korea: 1-5, Japan: 3-10, Macau: 6-12
14	2003	TIMSS	Fourth graders	28	Hong Kong, Japan, Singapore and Taiwan	13	Singapore: 1, Hong Kong: 2, Japan: 3, Taiwan: 4

Table 1.1 Continued

	Year	Survey Name	Target Population	N Edu. Systems	Participating East Asian Nations	Rank of the U.S.	Rank of East Asian Nations
15	2003	TIMSS	Eighth graders	50	Hong Kong, Japan, Korea, Singapore and Taiwan	18	Singapore: 1, Korea: 2, Hong Kong: 3, Taiwan: 4, Japan: 5
16	2006	PISA	15-year-olds	57	Hong Kong, Japan, Korea, Macao-China and Taiwan	32-36	Taiwan: 1-4, Hong Kong: 1-4, Korea: 1-4, Macao: 7-11, Japan: 6-13
17	2007	TIMSS	Fourth graders	43	Hong Kong, Japan, Singapore and Taiwan	13	Hong Kong: 1, Singapore: 2, Taiwan: 3, Japan: 4
18	2007	TIMSS	Eighth graders	56	Hong Kong, Japan, Korea, Singapore and Taiwan	13	Taiwan: 1, Korea: 2, Singapore: 3, Hong Kong: 4, Japan: 5
19	2009	PISA	15-year-olds	65	Hong Kong, Japan, Korea, Macao-China, Shanghai-China, Singapore and Taiwan	24-36	Shanghai: 1, Singapore: 2, Hong Kong: 3-4, Korea: 4-8, Taiwan: 4-8, Japan: 7-12, Macao: 9-12

a Mean achievement scores for the FIMS, SIMS and FIAEP comes from Medrick and Griffith 1992. Following PISA convention, I report the range of possible ranks for each educational system based on the 95% confidence interval of mean scores.

b PISA+, or the PISA 2000 study, was administered in 10 additional countries in 2002.

c Starting in 2003, PISA began assigning an upper and lower rank to each participating country. I have followed their convention here.

WHY THE ACHIEVEMENT GAP MATTERS

While success on the global stage might be linked to heightened national pride and unity, there are other reasons why we should be concerned about the mediocre performance of U.S. students on international assessments. Today's youth are competing with their counterparts in nations around the world for employment opportunities, and many jobs now require more math knowledge than the majority of U.S. students typically receive from a K-12 public school education (Dayal 2012; Duhigg and Bradshaw 2012; Flamm 2012). The New York *Times* reports that China proved a more advantageous setting for the manufacturing of Apple's iPhone in part because analysts at Apple had projected that it would take up to nine months to fill the 8,700 or so mid-level engineering positions the plant required in the U.S., compared to the 15 days it took to fill the same number of positions in China (Duhigg and Bradshaw 2012). The fact that U.S. students tend to possess weaker math skills than their counterparts around the world does not bode well for their future life chances in the global economy.

Additionally, research shows that international assessments serve as barometers of economic vitality. Using 40 years worth of achievement data from 50 nations to predict the average annual growth rate in GDP per capita while taking into account the security of property rights and openness to international trade in each nation over the same period, Hanushek and colleagues (2008) show that an increase of 50 points in achievement boosts a nation's annual economic growth rate by 0.63 percentage points.[2] To put that figure in context: if by 2000 the U.S. had reached the 1990 goal set by President Bush to become No. 1 in math and science achievement and effectively close the gap between the U.S. and East Asia, then the GDP in 2015 would be 4.5 percent greater than it would be without any achievement gains (Hanushek et al.

[2] Hanushek et al. explain that GDP growth in the U.S. has persisted despite our lackluster scores due to the strength of our tertiary education and our free market trading policies.

2008). That increase in GDP growth is not insignificant. According to Hanushek et al.'s calculations, the 4.5 percent boost is equivalent to what the nation spends yearly on K-12 education. In other words, if we figured out how to close the East Asian-U.S. achievement gap, then the investment would pay for itself.

THE COST OF FAILURE

International assessments might be the one place on the global stage where the U.S. does not excel. Some—including many East Asian-U.S. comparative education scholars—might argue that this is because, compared to sports and warfare, the U.S. cares relatively little about academics. But the amount of effort and money the nation has put into education over the last thirty years suggests otherwise. Educational reform has been a centerpiece of both Republican and Democratic Presidential Administrations since the 1980s. Per pupil spending for primary and secondary public education in the U.S. ballooned by 91.5 percent from 1980 to 2009 (the latest year with available data).[3] The U.S. now spends more per pupil on primary education than any East Asian nation under investigation. In 2008, the latest year for which data from all nations is available, the U.S. outspent Japan, Hong Kong, Singapore and Taiwan by at least 29 percent and as much as 263 percent per student.[4] Both the reform efforts as well as the increased funding suggest that the U.S. recognizes the importance of education and wants to improve the quality of education youth here receive.

---

[3] Data obtained from Table 191 of the 2011 *Digest of Education Statistics* published by the National Center for Education Statistics. Calculations based on the total expenditure per pupil in average daily attendance, in constant 2009-10 dollars, from 1980-01 ($6,861) and 2008-09 ($13,141).

[4] Data on per pupil spending on primary education in Hong Kong, Japan, Korea, Singapore, Taiwan and the U.S. come from the World Development Indicators database. Data from Taiwan were retrieved online from the Republic of China's Ministry of Education website. According to my calculations, the U.S. spent $10,531 per student on primary school education in 2008 compared to $8,157 in Japan, $3,908 in Hong Kong, $3,393 in Taiwan and $2,903 in Singapore.

However, neither reforms nor bigger budgets for education appear to have done much to boost student achievement in math and science. As Table 1.1 indicates, the nation's standing on international assessments has not changed much in the last few decades. On the TIMSS, U.S. eighth graders' scores increased by just 1.2 percent in math and 1 percent in science from 1999 to 2007. Among fourth graders, gains in math achievement were not much better, with the national average math score increasing 2.1 percent from 2003 to 2007 while the average science score was nearly stagnant, increasing just 0.6 percent over the same time period. The situation is similar if we use domestic assessments to measure gains in student achievement. Over roughly the same time period when per pupil spending on primary school education increased by 91.5 percent in the U.S., average 9-year-old test scores in math and reading on the National Assessment of Educational Progress (NAEP) only rose by 11 percent and 2.3 percent, respectively.[5]

The fact that the U.S. has tried to increase student achievement, and has spared no expense in the process, suggests that the problem stems not from a lack of concern but from a lack of understanding of what to fix.

The Value of International Comparative Research

To design educational reforms that work, educators and policymakers need to know why the U.S. is unable to provide all students with high quality learning opportunities. International comparative research can help educators and policymakers accomplish this task. By comparing the educational inputs provided in the U.S. to those of high-achieving nations, we can begin to

[5] Data on achievement scores come from the National Center for Educational Statistics (NCES) 2011 *Digest of Education Statistics* Table 125 (for reading) and Table 141 (for math). Scores on the NAEP tests in reading and math range from 0-500; in 1980 the average reading score of 9-year-olds was 215, by 2008 the score had risen by 5 points to 220. In 1982 the average math score of 9-year-olds was 219; by 2008 that score had risen by 24 points to 243 in 2008.

pinpoint the strengths and weaknesses in our educational system. Given the track record of Hong Kong, Japan, Singapore and Taiwan on international assessments, there are no better educational systems to use to measure the relative merits of the educational inputs U.S. students receive.

INFLUENTIAL RESEARCH IS OUTDATED AND INADEQUATE

East Asian-U.S. comparative research has informed and influenced U.S. educational policy in the past. Yet the last major study to provide an in-depth examination of the reasons for the disparity in achievement between multiple East Asian nations and the U.S., *The Learning Gap: Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education*, was published two decades ago. Advances in both data collection and methodology have rendered this work outdated and inadequate. However, its continuing influence on the field necessitates a brief summary. *The Learning Gap,* which was written by psychologists Harold Stevenson and James Stigler, represents the culmination of studies published by Stevenson and colleagues between 1982 and 1991 based on nonrepresentative data they collected during the 1980s, prior to the founding of PISA and TIMSS.[6]

The data for *The Learning Gap* were collected from five large, cross-cultural studies involving elementary school students, parents and teachers in Chicago, IL; Minneapolis, MN; Beijing, China; Sendai, Japan and Taipei, Taiwan that Stevenson undertook with support from the National Institute of Mental Health, the National Science Foundation and the William T. Grant Foundation. In collaboration with researchers in Japan, Taiwan and the U.S. Stevenson

[6] Stevenson and Stigler (1992) provide a full list of the studies on which *The Learning Gap* is based in the book's final section.

and his team devised achievement tests in reading and math that were free of cultural biases and contained grade-appropriate topics and concepts in math and reading.[7]

The results of Stevenson's math tests act as a precursor to the internationally comparable assessments: the math scores of American students at every grade level fall well below those of their East Asian counterparts. The primary focus of *The Learning Gap* is to search out the grounds for this disparity. Stevenson and Stigler split their attention between cross-regional differences in cultural beliefs regarding child-rearing and the importance of hard work to academic achievement, and cross-regional discrepancies in educational practices and institutional arrangements, such as teaching methods and the way the school day is organized. They identify a number of cross-regional differences in these two areas from their observations of classrooms and from their interviews with teachers and parents, and provide some reasonable explanations why these differences might attribute to the achievement gap.

The main thrust of the book, and the dozens of articles published by Stevenson and colleagues before and since, appears to be that East Asians excel in math due to the cultural values of Confucianism, which emphasize education and success through hard work. The poor performance of U.S. students, on the other hand, reflects a pattern of cultural values and beliefs in the U.S. that does not support math achievement. In this theory, cultural values account not only for cross-regional differences in student achievement but also for cross-regional variation in school practices and educational policies. In other words, the cultural values embodied in

[7] The first study was administered in 1980 to roughly 240 students in each of three grades—kindergarten, first and fifth—from 10 elementary schools in Minneapolis, Sendai and Taipei (Stevenson et al. 1986). The team conducted a follow-up study in 1984 of the first-grade students who had participated in the 1980 study and were now in the fifth grade. The next two studies took place in 1986. One includes first and fifth graders and their parents in Sendai, Taipei and Chicago. The other targeted first, third and fifth graders, their parents and teachers in Beijing and Chicago. The final study took place in 1990 and targeted fifth graders in Minneapolis, Sendai and Taipei.

Confucianism gives rise to more effective and efficient educational systems as well as to citizens who invest more time and effort into scholastic achievements.

The studies on which *The Learning Gap* is based garnered significant media attention when they were first published; Stevenson and Stigler embarked on a national speaking tour soon after the book's release. Despite its age, *The Learning Gap* remains in print in the 1994 paperback edition. It represents the touchstone for research on East Asian academic success and math achievement. A cited reference search on the ISI Web of Science shows that it has been referenced in over 400 published works. The authors' opinions continue to shape understanding of East Asian education; the book has been included in the syllabi for dozens of college courses throughout the past decade. While both Stevenson and Stigler separately have tinkered with the conclusions around the edges, and some attempts have been made to clarify the gap between Japan and the U.S. (LeTendre 1999) or between Chinese and U.S. teachers (e.g. Ma 1999), *The Learning Gap* represents the last comprehensive study that integrated data from more than one East Asian nation to assess the root causes of the East Asian-U.S. math achievement gap.

Despite its popularity and influence, the message of *The Learning Gap* is not entirely clear. This is reflected in the book's reviews. One reviewer suggests that the focus is on what students, teachers and parents do at home and at school that differs fundamentally between East Asia and the U.S. (Yano 1993), while another reviewer claims that the emphasis is on attitudinal, philosophical and behavioral differences between two cultures (Pierce 1993). Despite Stevenson and Stigler's dual focus on cultural beliefs and formal educational arrangements, their understanding of the relationship between the two is unresolved. It seems to be that cultural values influence the institutional arrangements of the educational system and continue to shape the day-to-day activities of students, parents and teachers. However, the remedies the authors provide to increase student achievement in the U.S. include not only changes to cultural values (example: “believe in effort”) but also specific policy recommendations (such as national curriculum standards and more time for teachers to lesson plan).Yet it is unclear why they believe the educational reforms they recommend would be successful without a concurrent shift in cultural values.

Another issue is that Stevenson and Stigler do not distinguish between assumptions and facts in their conclusions and policy recommendations. For example, Stevenson and Stigler assume that teachers in East Asia are more effective than their counterparts in the U.S. (the topic I empirically examine in Chapter 3), and focus their attention on the belief system, teaching methods and teacher training requirements in Japan and Taiwan for explanations and policy recommendations. Similarly, as I discuss in Chapter 2, they assume that the uniformity of instruction they witnessed across classrooms in Japan and Taiwan is the result of central control of the curriculum by the Ministries of Education in those nations and prescribe national curriculum standards as a possible solution for the U.S. However, as I explain in Chapter 2, the information compiled by TIMSS casts doubt on this claim. Despite the lack of analytic rigor, many of their arguments concerning East Asian education and beliefs have become taken-for-granted facts in the literature on East Asian education. Until we can separate assumptions from facts, policy recommendations seem premature.

The field of East Asian-U.S. comparative educational research by and large continues to be mired in empirical and methodological weaknesses. Nationally representative data are still relatively underutilized. While conducting background research for the chapter on parents, for example, I found that the most well-cited works tends to rely on data from East Asian immigrants to the U.S. (for example: Kao 1995; Okagaki and Frensch 1998; Schneider and Lee

1990; Sun 1998). Given that immigrant Asian parents are raising their children in the same environment as their U.S.-born, white counterparts, it is no surprise that scholars attribute the greater educational investments of immigrant Asian parents (compared to U.S.-born white parents) to cultural values as opposed to contextual factors. Some immigrant scholars have moved beyond an emphasis on East Asian culture to suggest that macro sociocultural factors also motivate immigrant Asian parents' involvement. For example, Sue and Okazaki (1990) contend that discrimination against Asians in many sectors of the U.S. economy prompts immigrant Asians to view education as their primary vehicle to occupational attainment. Yet a reliance on data from immigrant Asians to the U.S. precludes many scholars from asking how features of schools in East Asia might also contribute to what appear to be greater investments in education among East Asian parents.

Schools in East Asia would influence how immigrant Asian parents support their children's education if immigrant Asian parents factor into their decisions their past experiences as students in East Asian educational systems. If this happens, then we would expect to find some differences in parental involvement between Asian parents who were born in the U.S. and their counterparts who were born and raised in Asia. Indeed, both Kao (2004), who analyzed data from the National Education Longitudinal Study 1988 (NELS), and Pong, Hao and Gardner (2005), who relied on data from the National Longitudinal Study of Adolescent Health, have shown statistically significant differences in the parenting practices of Asian-born Asian parents and U.S.-born Asian parents. More evidence that immigrant Asian parents' decisions regarding parental involvement are influenced by their past experiences in East Asian schools comes from Louie's (2001) ethnographic study of immigrant Taiwanese parents and their college-aged children in the New York City area. Louie (2001) notes that one of her respondents, a Taiwanese mother whose son was unorganized and acting out when he was in elementary school, used a notebook to communicate with her son's teacher so that she could help her son stay on track. While Louie does not expressly mention this fact, Taiwanese elementary schools employ the same technique to keep parents in the loop concerning their children's homework and to provide parents with updates on their children's progress (Stevenson and Stigler 1992). It seems highly likely that Louie's respondent got the idea for the communication notebook from her own experiences in a Taiwanese elementary school.

Problems arise when scholars equate immigrant Asian parents or students with parents or students in East Asia. This is mainly for two reasons. First, the context in which immigrant Asians parents raise their children is different from the context in which parents in East Asia raise their children. Similarly, the context in which immigrant Asian students find themselves is very different from the context in which students in East Asia find themselves. The social context almost certainly influences what parents decide to do to support their children's education (a point I elaborate on in Chapter 4)—and how students navigate their way through the school system. Second, and most important, immigrant Asians to the U.S. are a select group (e.g. a limited few who chose to leave their native country to live and work in the U.S.), and tend to be more educated than their native countrymen (Feliciano 2005). For these reasons, immigrant Asians simply are not representative of their native populations. Unfortunately, it is not uncommon for researchers to generalize their findings from immigrant Asians to the Asian population in Asia, with the result being that research rarely considers the influence of school and community factors on parenting styles or student performance and, in turn, on the East Asian-U.S. achievement gap.

Beyond these empirical and methodological flaws, the guiding theory is also weak. Prevailing wisdom suggests that cultural values propel East Asian nations to the top of the charts in math achievement. First, the theory implies that U.S. culture is inferior to East Asian culture. It is a transparently defeatist argument: without a seismic cultural shift, math achievement will always be higher in East Asia than in the U.S. This kind of cultural depravity model has been replaced in sociology by theories that explore both structural and cultural factors to understand achievement gaps between groups of students within the U.S.

The guiding theory to explain the East Asian-U.S. math achievement gap is also flawed because it neglects the diversity of cultural legacies within East Asia. Hong Kong, for example, while part of East Asia, was ruled by the British for more than 100 years. All spheres of life in Hong Kong were thoroughly influenced by this concession, including economic, educational, political and cultural. Singaporean culture, too, is influenced by more than Confucianism. The nation-state also underwent a period of British rule during the 19th century. While the majority of its citizens today are of Chinese descent, sizeable portions of the population are also Malay and Indian and almost 30 percent of the population is foreign-born. Thus, Singapore includes a number of different ethnic and religious groups, which contribute to the plurality of the nation-state's cultural legacies. Given the paramount importance of cultural legacies to the prevailing theory on the East Asian-U.S. achievement gap in math, it is particularly troubling that the cultural diversity within East Asia is so often overlooked.

Lastly, the prevailing theory on the East Asian-U.S. achievement gap in math might be faulted for failing to acknowledge the disconnect between values and actions, or between what individuals believe and what they do. While cultural legacies, or macro-level value systems, can shape micro-level actions and behaviors, the two are not always in sync. The theoretical one-to-one correspondence between values and actions more often than not leads to the conclusion that nothing can be done to eradicate the gap.

IMPROVING RESEARCH ON THE EAST ASIAN-U.S. MATH ACHIEVEMENT GAP

A sociological perspective would bring much needed clarity to this research. Sociologists of education who follow James Coleman tend to view individual schools as mostly closed social systems, which means that students' actions and behaviors are only guided by the norms and expectations of the adult community to the extent that school staff, parents and community members work to instill these values in students.[8] Explaining the chief reason why he chose to study high schools (for *The Adolescent Society*), Coleman (1994: 31) wrote:

> ... high schools seemed to be one of the few social contexts in modern society that constituted largely self-contained social systems. Most adolescents directed their attention inward. Status among the adolescents in the school was more important than status outside. The youth could not easily leave the system and choose a different one. This meant that the processes which generate norms, systems of status, cleavage and conflict, in short the community's functioning, were first of all internally generated (though with influences from school staff, parents, and community), and second, they were effective in shaping the behavior of the members of the system.

In this way, the higher average math achievement of students in East Asia compared to U.S. students can be seen as the product of educational systems with high achievement norms *and* with school staff, parents and community members who work to effectively instill those norms in youth. The implication is that a cultural legacy that values education *on its own* is not enough to propel a nation to the top of international assessments in math. Educational systems must endow educators with the resources and skills they need to foster high achievement norms among students. Given that parents also play an outsized role in their children's early educational

[8] Whether schools are viewed as primarily open or closed social systems is a matter of debate within the sociology of education. While Coleman clearly viewed schools as closed systems, Chubb and Moe (1997: 364), for example, view schools as open systems "and thus products of their environments." For more insight on Coleman's understanding of schools as open social systems see Schneider (2000).

development, how educational systems engage parents in the schooling process (either directly, through websites or brochures, or indirectly, by preparing educators to work with parents) might also impact the extent to which schools can support and maintain high achievement norms among students.

Distinguishing schools as essentially self-contained social systems allows us to view cultural legacies as values that are instilled in youth as opposed to beliefs that are inherited through blood rights or birthplace. Moreover, this perspective explains why it is beneficial to examine cross-regional differences between East Asian nations and the U.S. in formal educational arrangements and everyday practices as the former shape and the latter shed light on educators' capacities to encourage achievement. However, American sociologists have not made significant use of the internationally comparative data that have been collected in the decades since Stevenson and colleagues conducted their studies. This is unfortunate because—even though Stevenson and Stigler (1992) identify a number of cross-regional differences in formal educational arrangements and everyday practices—the evidence they provide to support many of the factors that they pinpoint as causes of the achievement gap loses some appeal upon closer inspection. For example, as I detail in Chapter 4, Stevenson and colleagues argue that the achievement gap exists in part because compared to U.S. parents, East Asian parents are more effectively involved in their children's education. But their measures of effective involvement do not hold up to empirical scrutiny.

For these reasons, the aim of this dissertation is to put research on the East Asian-U.S. achievement gap on sounder empirical and methodological and theoretical footing. To accomplish this task, I focus on just three of the many variables that Stevenson and Stigler covered—the uniformity of math instruction provided, teacher effectiveness and parental involvement—and rely on nationally representative data from the 2003 and 2007 TIMSS fourth-grade population. Even limited to three variables and using two rounds of TIMSS, there are not enough data to make any concrete conclusions. However, putting the research on a sounder empirical and methodological basis will help to shine the spotlight on its strengths and weaknesses. Further, in the future when more complete data become available, additional research will provide a foundation for policy recommendations.

DIVERSITY OF EAST ASIAN EDUCATIONAL SYSTEMS

The majority of data that Stevenson and Stigler (1992) collected from East Asia come from Japan and Taiwan. This dissertation relies on data from all East Asian nations that participated in the TIMSS fourth-grade populations: Japan, Singapore and Taiwan. By expanding the number of East Asian nations in the analysis, this dissertation throws into relief the diversity of educational institutions and practices within East Asia. In terms of elementary school assignment, for example, in Japan and Taiwan the schools children attend are entirely determined by home address, while in Hong Kong and Singapore parents are given some choice as to their children's schools. There are also differences across nations in the methods used to formally sort students by ability. In Japan and Taiwan, high-school entrance examinations, which are taken in the grade 9, are used to sort students into academic or vocational high schools of varying levels of prestige. In Singapore, on the other hand, students in grade 6 take a Primary School Leaving Examination, which determines their track placement. The system in Hong Kong is similar to Singapore in that students are allocated to different positions in the system depending on their performance on three examinations taken in grade 5 and 6. Additionally, while all East Asian nations have national curriculum standards, the methods they use to implement those standards, and the amount of autonomy schools have over textbooks and curriculum, varies within the region.

The fact that there is not a monolithic educational system in East Asia suggests that following the typical route of international comparative sociological research on education and focusing on institutional arrangements of educational systems will not bring us closer to understanding why the East Asian-U.S. achievement gap in math exists. Nor does it seem appropriate to continue to vitiate East Asian-U.S. comparative research by ignoring the diversity of institutional arrangements within East Asian educational systems. Instead, this dissertation treats each East Asian nation as a distinct educational system to assess the reasons for the gap.

DISSERTATION OVERVIEW

What is also missing from research on the East Asian-U.S. achievement gap is a rigorous quantitative analysis of high-quality internationally comparative data. This dissertation utilizes data from the 2003 and 2007 TIMSS fourth-grade population to build on Stevenson and Stigler's (1992) investigation of the reasons for the math achievement gap between elementary school students in the U.S. and East Asia. Stevenson and Stigler's review of cross-regional differences in the educational infrastructure of China, Japan, Taiwan and the U.S. ranges broadly from the number of hours that teachers are scheduled to teach each week to who is responsible for sweeping and cleaning the classroom floors. To begin to put research on a sounder empirical and methodological footing, I narrow the focus of this dissertation to three major influences on student achievement—curriculum, teachers and parents.

In the remaining sections of this chapter, I discuss the TIMSS data, the methods that are common to all chapters, and the magnitude of the math achievement gap between fourth-grade students in East Asia and the U.S., according to TIMSS. In Chapter 2, I turn attention to the implemented curriculum, or the quantity of math instruction that students receive. Here, I empirically assess Stevenson and Stigler's (1992) assertion that math instruction is more uniformly distributed across elementary-school classrooms in East Asian educational systems than in the U.S., and address the question of whether average math achievement is higher in East Asia because of this. I build on the discussion of curriculum in Chapter 2 by examining how the quantity of instruction teachers provide influences student achievement as well as measures of teacher effectiveness. In so doing, I empirically assess Sørensen's (with Hallinan 1977; with Morgan 2000) argument that student learning depend on how much students have been taught. I also investigate the possibility that the East Asian-U.S. achievement gap exists, as Westbury (1992) contends, not because U.S. teachers are less effective than their East Asian counterparts, but because East Asian teachers tend to teach more TIMSS math topics than U.S. teachers.

In Chapter 4 I shift focus to explore the parental involvement gap between East Asia and the U.S., which Stevenson and colleagues finger as a prime cause of the achievement gap. The prevailing theory to explain why levels of parental involvement are supposedly higher in East Asian nations than in the U.S. is that Confucianism compels East Asian parents to make their children's education their top priority. This cultural deprivation model is inherently defeatist and leads scholars too often to conclude that nothing can be done to eradicate the East Asian-U.S. parental involvement gap. I draw on case studies of Japanese preschools and elementary schools that have been published since Stevenson and Stigler (1992), as well as sociological theory, to put forth an alternative argument that schools can make a difference in the proportion of parents who are involved in their children's education. Finally, in Chapter 5, I bring together the findings from each chapter, and discuss the broader methodological and theoretical implications of this work.

DATA & METHODS

TIMSS was developed by the International Association for the Evaluation of Educational Achievement (IEA) and has been carried out every four years since 1995 to collect data on math and science achievement in educational systems around the world. The most recent rounds of TIMSS have two target populations: the grade with the largest proportion of 9-year-olds, which is commonly referred to as the fourth-grade population, and the grade with the largest proportion of 13-year-olds, which is commonly referred to as the eighth-grade population.[9] For this dissertation, I analyze data from the 2003 and 2007 rounds of the TIMSS fourth-grade population. I limit the analyses to the more recent rounds as changes to the questionnaires since 1995 make comparisons with the earlier round difficult.

TIMSS is unique among internationally comparable datasets in using curriculum, broadly defined, as its major organizing concept. TIMSS conceptualizes curriculum in three forms: 1) the intended curriculum, which represents the topics that governing bodies intend for students to learn, 2) the implemented curriculum, or the topics that are actually taught in the classroom, and 3) the attained curriculum, or the material that students have learned, which is gauged by their achievement on the TIMSS assessments. Additionally, TIMSS asks students, teachers and school principals to fill out questionnaires regarding the contexts for learning math and science. This organizing structure has important implications for the type of data TIMSS gathers and for the content of its assessments. The end result is an internationally comparable dataset that provides

[9] Since 2003, TIMSS has been conducted at two grade levels. In 1999, just the eighth-grade population was targeted. The 1995 round targeted students in the third, fourth, seventh, eighth and twelfth grades.

more information than any other on the quantity of instruction that students receive, the extent to which the quantity of instruction that students receive typically varies within nations, and how much students in each nation have learned of what they have been taught.

Unlike PISA, which focuses on assessing students' abilities to apply their knowledge and skills to real-life challenges, TIMSS aims to measure the extent to which students have mastered the topics and skills that they have been taught in the classroom. Extensive consultations among representatives from each participating nation, national committees and an international panel of math and science education and testing experts are conducted before each survey round to ensure that the individual math and science curricula topics included in the assessments are appropriate for each grade level in each participating nation. While the TIMSS assessments do not consist solely of the topics and skills included in the curriculum of all participating nations, the extensive care that goes into determining their content ensures that they are not "general use" assessments of students' aptitude (such as the PISA or the Stanford Achievement Test). Instead, the TIMSS assessments measure the outcome of the principal activity of schools, which is what students have been taught in the classroom. In this way, TIMSS offers the best kind of assessment for studying the influence of teachers on student achievement.

The TIMSS assessments are administered near the end of the school year in each nation and contain questions in both constructed response and multiple-choice formats. The number of assessment items included in the entire test is too large for any one student to complete. For example, it would take an estimated 8.5 hours for a fourth-grade student to finish the 2007 TIMSS assessment (Mullis et al. 2005: 99). Therefore, TIMSS employs the matrix-sampling technique typically used for large-scale assessments to distribute assessment items across student test booklets such that each booklet contains a balanced subset of the math and science items.

The assessment time for individual students was 72 minutes in 2003 and 2007, although the assessment design was modified slightly in 2007 to increase the amount of time students had to complete each section of the assessment. To derive the score students would have received had they completed the entire assessment, TIMSS uses Item Response Theory (IRT) scaling with conditioning and multiple imputation (for a detailed account of this method see Gonzalez et al. 2004). IRT was originally developed by the Educational Testing Service and has been a popular technique for use on large-scale surveys since the 1970s (Gonzalez et al. 2004). To correct for the error inherent in the imputation process, TIMSS provides five plausible values of math and science achievement for each student. Achievement scores are internationally scaled to have a mean of 500 and a standard deviation of 100. For accuracy, I conduct each analysis of student achievement five times, once on each plausible value, and combine estimates using Rubin's (1987) combination method.

According to Rubin (1987), a single set of results can be obtained from a data analysis performed *n* times (in this dissertation, *n* = 5) by saving the estimated coefficients and standard errors from each trial and combining them in the following manner. For simplicity, assume that each trial yields a coefficient $\hat{X}_t$ and a standard error $\widehat{U}_t$ with *t*=1, 2... *n*. For coefficients, the formula for the combined estimate is the average of $\hat{X}_t$ across *n* trials is:

$$X = \frac{1}{n}\sum_{t=1}^{n}\hat{X}_t$$

There are four steps in the calculation of standard errors. The first is to compute the within-imputation variance, *W*, which is the average of $\widehat{U}_t^2$ across *n* trials:

$$W = \frac{1}{n}\sum_{t=1}^{n}\widehat{U}_t^2$$

The next step is to calculate the between-imputation variance, B, which is the sum of the squared difference between the estimated coefficients $\hat{X}_t$, and the overall coefficient X divided by n-1:

$$B = \frac{1}{n-1}\sum_{t=1}^{n}(\hat{X}_t - X)^2$$

Next, the total variance, T, is obtained using the following formula:

$$T = W + \left(1 + \frac{1}{n}\right)B$$

Lastly, to calculate the overall standard error requires taking the square root of the total variance, T.

TIMSS created a simple and easy sampling design, which National Research Coordinators (NRC) in each nation could implement with limited resources, that yields accurate and efficient samples of schools and students. TIMSS ensures that NRCs use the correct sampling procedures by providing easy-to-use manuals, software and expert assistance from the IEA Data Processing Center and Statistics Canada. NRCs were free to adapt the sampling design for their educational systems, but all adaptations had to be approved by the TIMSS International Study Center at Boston College (Foy and Joncas 2004).

Given the focus on curriculum, in TIMSS both the classroom and the school—in addition to students—are potential units of analysis. Therefore, the sampling design takes into account all three—schools, classrooms and students—in order to achieve sampling precision and data quality at all three levels. TIMSS relies on a two-stage stratified cluster sampling design. The first stage consists of schools that enroll the target population. Nations could decide to stratify schools (by factors such as geographic region or school type) to improve the efficiency of the sample design, to ensure adequate representation of specific subgroups of students in the sample, or to oversample certain subgroups of schools (Foy and Joncas 2004). Schools are then selected, along with a replacement school, with probability proportional to size from each stratum. In the second stage, one or more intact math classrooms from the target grade in each selected school are sampled.

While TIMSS takes into account specific features of each nation's educational system to determine the number of schools and students that each has to sample to yield reliable estimates, participating nations are required to sample a minimum of 4,000 students and 150 schools. The end results are probability samples that give accurate weighted estimates of population parameters. TIMSS provides sampling weights for each unit of analysis (students, classrooms and schools) that reflect the inverse of the probability of selection, controlling for non-responding units (for a detailed account see Joncas 2004; Joncas 2008). TIMSS provides jackknife replicate-weight variables to use to estimate error variances. The jackknife repeated replication technique estimates sampling error through repeated re-sampling of the data. The variation of the technique used by TIMSS computes estimates once for the entire sample and once again for up to 74 "pseudo-replicate" samples of the original data. The jackknife variance estimation is the variance between the estimates for each of the replicate samples and the entire sample. In this dissertation, all descriptive statistics are computed using this jackknife technique.

Although it is generally nations that participate in TIMSS, there are a few exceptions. For example, Belgium, which has two educational systems (one French-speaking and one Flemish-speaking), participated in the 2003 round but only sampled students, classrooms and schools from the Flemish-speaking system. Additionally, TIMSS also permits individuals states or provinces within nations to take part in the study as "benchmarking participants." These participants are held to the same sampling standards as regular participants, with the exception that the minimum number of students and schools that must be sampled is lower. Three states have participated in either the 2003 and 2007 TIMSS; in 2003 Indiana participated with a sample of 2,236 students in 56 schools, and Massachusetts and Minnesota participated in 2007. (The Massachusetts data include 1,747 students from 47 schools. The Minnesota data include 1,846 students from 50 schools.) Although the number of states that have participated so far is tiny, and their sample sizes are small, where possible I include these data in my analyses.[10]

Hong Kong, Japan, Singapore and Taiwan and the U.S. participated in both the 2003 and 2007 rounds of TIMSS, so there is a substantial amount of data from each nation under investigation. The U.S. sample includes data from 17,725 students in 994 classrooms in 505 schools. For Hong Kong, there are data from 8,399 students in 292 classrooms in 258 schools. The Japan sample contains information from 9,022 students in 339 classrooms in 298 schools, while the Singaporean sample consists of data from 11,709 students in 536 classrooms in 359 schools. Finally, in the Taiwan sample, there are data from 8,792 students in 324 classrooms in 300 schools. Although this dissertation focuses primarily on these educational systems and the nationally representative sample from the U.S., I draw on data from other educational systems when the research question dictates. In Chapter 2, to assess the relationship at the system level between the uniformity of instruction across classrooms and student achievement, I rely on data from every educational system that participated in each survey round because the estimates become more reliable as the sample size increases. Additionally in Chapter 2, I draw on the state data because they allow me to examine whether math instruction is more uniformly provided within states than across states within the U.S. In Chapter 3, where the focus is to compare the efficacy of teachers in the U.S. and East Asia, I rely solely on data from those national samples.

[10] I do not merge the individual state data with the national sample in my analyses. The analyses of state data will be more fruitful once the 2011 TIMSS data are released because nine states participated in that round.

Finally, in Chapter 4, I use data from every high-achieving nation so I can put the proportion of parents in East Asian nations who are involved in their children's education into perspective.

THE EAST ASIAN-U.S. ACHIEVEMENT GAP

To see how well the U.S. as a whole and each of the three states that participated in either round of TIMSS instill math knowledge in fourth-grade students compared to educational systems in East Asia, I compare the average mathematics score of each educational system. Figure 1.1 graphs these results. Nations are listed in alphabetical order from left to right, and the three states are listed on the right, after the U.S. data. Results from the 2003 round are in black. In the 2003 round, Singapore, Hong Kong, Japan and Taiwan ranked first through fourth, in that order, in math achievement. The U.S. ranked 12th, and scored almost 50 points or one-half of a standard deviation below the two lowest-scoring East Asian nations of Japan and Taiwan and 76 points below the highest-achieving nation of Singapore. Indiana state outperformed the nation as a whole, averaging a score about 15 points higher than the national average. These trends continued in 2007. Hong Kong, Singapore, Taiwan and Japan ranked first through fourth, in that order, in 2007 while the U.S. ranked 11th. The U.S. as a whole averaged a score roughly 40 points below the lowest-scoring East Asian nation of Japan and 78 points below the highest-achieving nation of Hong Kong. The U.S. states that participated in the 2007 round, Massachusetts and Minnesota, both outperformed the nation as a whole. Massachusetts averaged an impressive score of 572, which puts its performance on par with Japan and Taiwan. The average math score of Minnesota is 25 points above the average for the U.S.

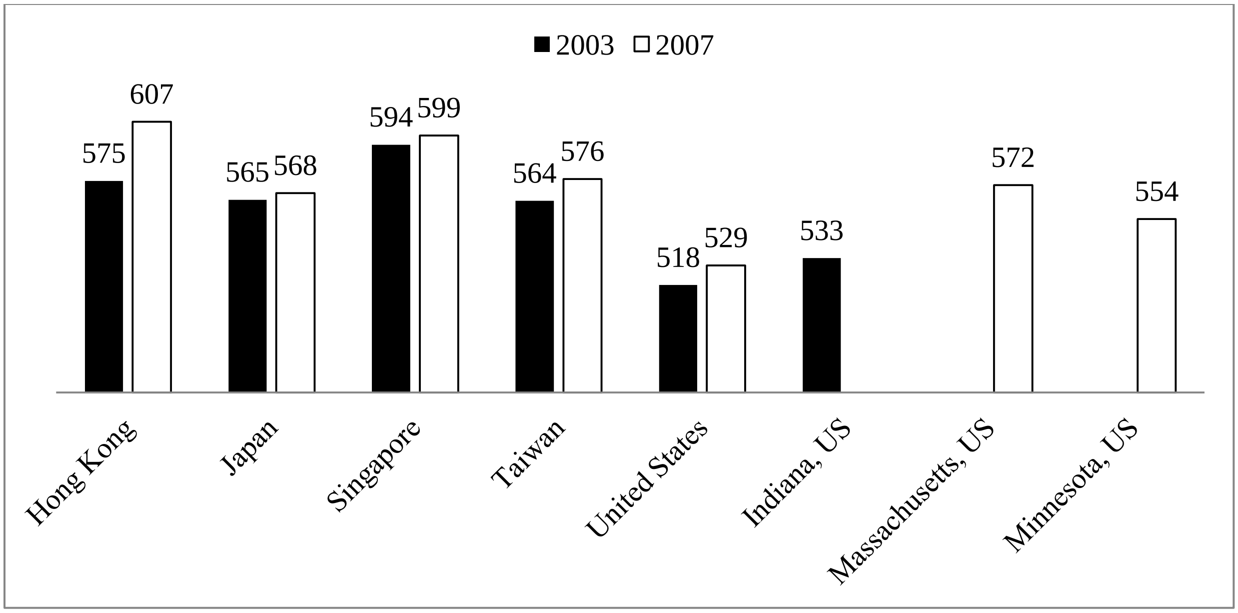

Figure 1.1 Fourth Grade Mean Math Achievement Scores on the 2003 and 2007 TIMSS

Another way to evaluate the capacity of educational systems to impart math knowledge is to compare the percentage of students in each educational system whose scores reached the advanced international benchmark. This provides insight into the proportion of highly accomplished math students each system is producing. The percentages of fourth-grade students from both rounds of TIMSS who reached the advanced international benchmark are displayed in Figure 1.2. In Figure 1.2, nations are listed in alphabetical order from left to right with the state data following the U.S. data on the right. The data from the 2003 round are displayed in solid black.

According to Figure 1.2, the percentage of students reaching the advanced international benchmark in math in the U.S. is less than half that of any East Asian nation in either round. In 2003, 7.4 percent of U.S. fourth graders reached the advanced benchmark compared to 37.8 percent in Singapore, 21.7 percent in Hong Kong, 20.6 percent in Japan and 16.0 percent in Taiwan. The percentage of highly accomplished math students from the Indiana sample, 7.5

percent, is on par with that of the U.S. In the 2007 round, 10.0 percent of U.S. students reached the advanced benchmark compared to 40.7 percent of fourth graders in Singapore, 40.4 percent in Hong Kong, 24.1 percent in Taiwan and 22.8 percent in Japan. On this metric, Massachusetts and Minnesota again outperformed the U.S. as a whole. The percentage of highly accomplished math students in the Massachusetts sample (22.5%) is on par with that of Japan, while the percentage from Minnesota (18.2%) is nearly double that of the U.S.

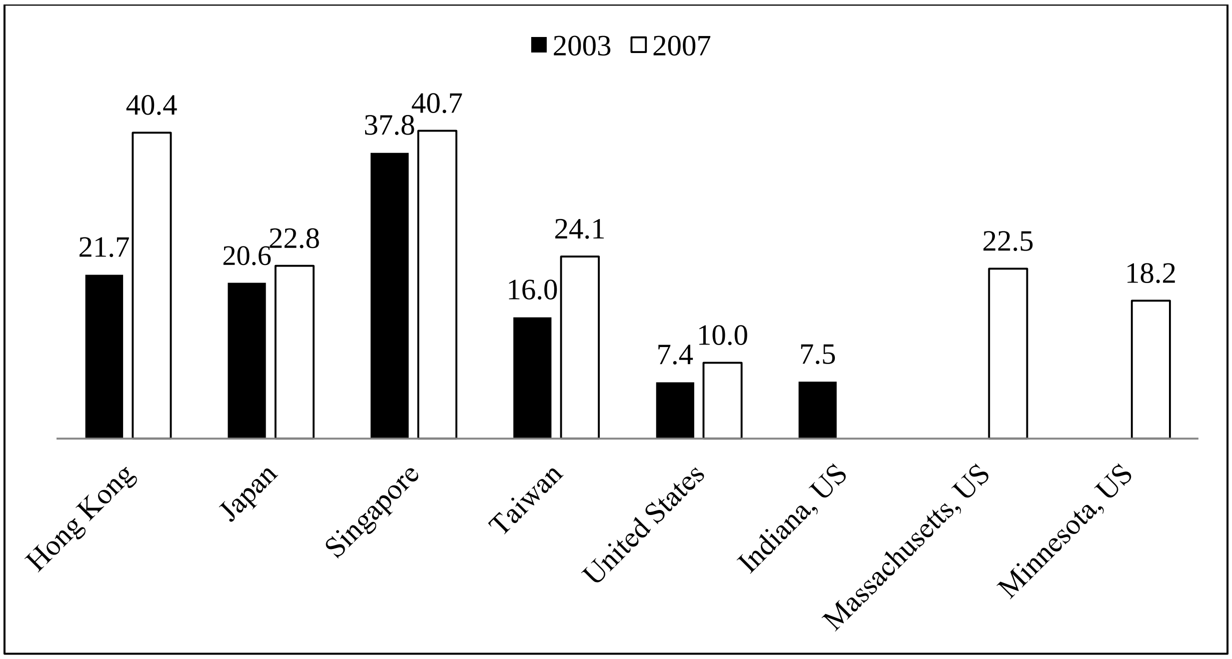

Figure 1.2 Percentage Meeting the Advanced International Benchmark in Math

Not only is the U.S. producing a smaller percentage of students who achieve at the highest level, but a look at the percentage of students who scored at or below the low international benchmark in each nation tells us that the nation is also producing a significantly larger percentage of students whose scores indicate that they do not have the ability to apply basic math knowledge in straightforward situations. Figure 1.3 provides a look at these data and is formatted in the same manner as the two previous figures. According to Figure 1.3, the percentage of low achieving math students in the U.S. in 2003 is almost three times that of Japan, the nation with the next-highest percentage of low-achieving math students. In that round, 28.3 percent of U.S. fourth graders scored at or below the low international benchmark in math, compared to just 6.4 percent of students in Hong Kong, 8.2 percent in Taiwan, 8.8 percent in Singapore and 11.2 percent in Japan. The percentage of students from Indiana whose scores fall at or below the low international benchmark, 18.1 percent, is lower than the U.S. sample, but still almost fifty percent higher than that of Japan.

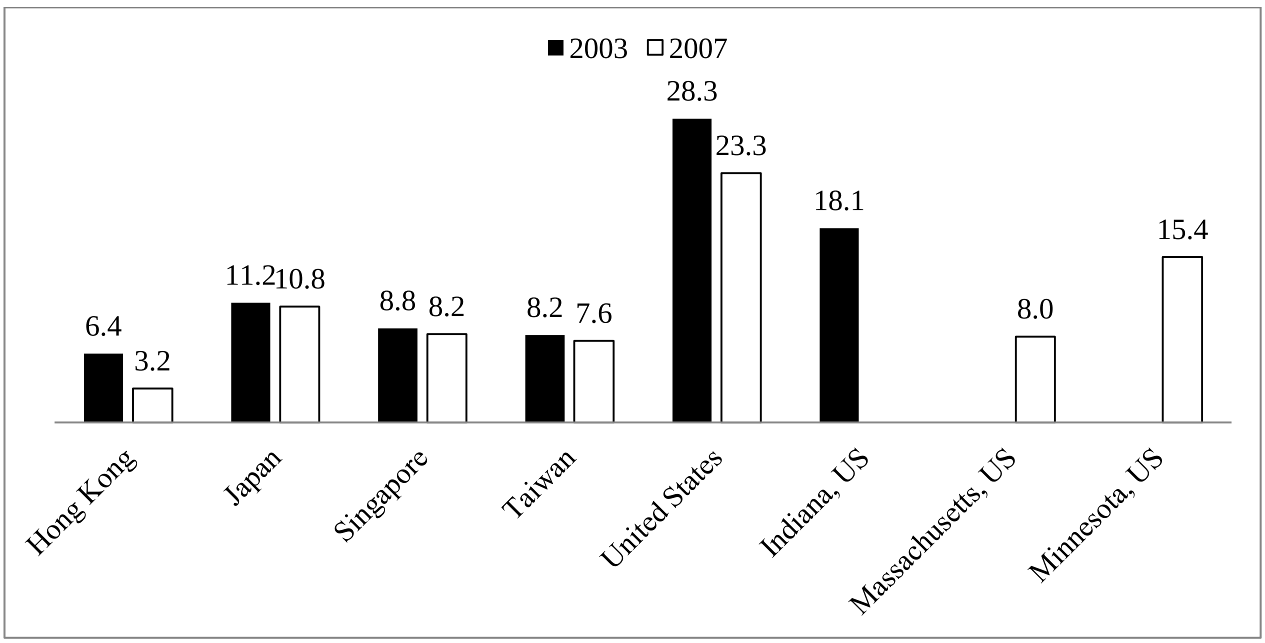

Figure 1.3 Percentage Scoring at or Below the Low International Benchmark in Math

In 2007, the U.S. sample again contained a sizeable proportion of low-achieving math students. Twenty-three percent of fourth-grade students in the U.S. failed to reach the intermediate benchmark in math, compared to only 3.2 percent in Hong Kong, 7.6 percent in

Taiwan, 8.2 percent in Singapore and 10.8 percent in Japan. The percentage of students whose scores fall at or below the low international benchmark in Minnesota is also higher than in any East Asian nation, at 15.4 percent. However, the percentage from the Massachusetts sample, 8.0 percent, is on par with Japan and Taiwan.

In sum, according to results from the TIMSS 2003 and 2007 math assessments, the U.S. as a whole continues to produce fewer elementary-school students with strong math skills than nations in East Asia. Compared to Hong Kong, Japan, Singapore and Taiwan, the U.S. averaged a significantly lower score on both math assessments. Further, the U.S. produces a greater percentage of low-achieving math students than any East Asian nation that participated in the TIMSS 2003 or 2007 fourth-grade population. However, the results from Massachusetts are promising. The data from Massachusetts show that it is possible for the U.S. to succeed at imparting math knowledge at a rate that is nearly indistinguishable from that of East Asian nations and merits further study. For now, however, we can learn a great deal from examining education in Hong Kong, Japan, Singapore and Taiwan—which have long succeeded in producing mathematically-adept youth and for which we have ample data—for insight that will help us improve the state of education in the U.S.

Chapter 2

Curriculum

MARVELING OVER THEIR VISITS to numerous fifth-grade classrooms during a short sojourn in Taiwan, Stevenson and Stigler (1992: 136) describe the experience as one of “déjà vu.” This is because, as they explain, “All fifth-graders within the school, within different schools in the city, even within all of Taiwan, are studying the same lesson.” Stevenson and Stigler are not alone in holding in high regard educational systems that have achieved significant uniformity of instruction across same-grade classrooms. In these educational systems, there is less variation in the quantity of instruction provided, more agreement among teachers regarding the curriculum, and less chance that some students do not receive training in basic knowledge and skills (Kerckhoff 1995; Stevenson and Baker 1991; Stevenson and Stigler 1992). In other words, educational systems that have achieved significant uniformity of instruction across classrooms appear to be meeting one of the core challenges of schooling, which is to provide equal educational opportunities to all students (Van de Werfhorst and Mijs 2010). This is not an easy challenge; the central aim of the last two Presidential Administrations has been to decrease the inequality of educational opportunities provided within the U.S. The abundance of scholarly and media critiques of President George W. Bush’s No Child Left Behind policy and

President Barack Obama's Race to the Top funding initiative attest to the ongoing nature of the debate on the policies and practices that would best accomplish this task. Despite the importance of providing equal educational opportunities, international comparative research rarely pays attention to the relative uniformity of instruction offered within educational systems.

There is good reason to believe that realizing significant uniformity of instruction across same-grade classrooms is imperative, not only for equality but also for student achievement. This is because systems that fail to educate large portions of their youth populations should have lower average achievement scores than systems that tend to provide all students with basic knowledge and skills. Stevenson and Stigler (1992) suggest that the relative lack of uniformity of instruction across classrooms in the U.S. is one reason for the East Asian-U.S. achievement gap. However, the authors do not empirically examine this hypothesis. Instead, they contend that the uniformity of instruction across classrooms in Japan and Taiwan stems from the central control of the curriculum by the Ministry of Education in both nations, and prescribe national curriculum standards as a possible solution for the U.S. However, there is very little reason to believe that national curriculum standards *on their own* encourage either uniformity of instruction or student achievement.

The TIMSS data call into question the hypothesized link between national curriculum standards and student achievement. First, only 7 of the 42 nations that participated in the 2003 or 2007 fourth-grade population do *not* have national curriculum standards in math. Of those seven nations (Australia, Belgium, Canada, Colombia, Germany, Iran and the U.S.), three (Belgium, Germany and Iran) have officially defined national educational standards.[11] Further, the

[11] The Flemish-speaking educational system in Belgium determines the curriculum for exit examinations that have consequences for students and therefore influence the curriculum used in schools; In Germany, the sixteen states determine their own curriculum but since 2004 each state has committed to implementing the national educational standards; On the 2007 survey, Iran responded that it does not have national curriculum standards but wrote in the

correlation between national curriculum standards (educational system has national curriculum standards=1; if not=0) and average achievement in math is small and negative (-0.05 for the 2003 data and -0.03 for the 2007 data), suggesting that the relationship is weak and in the opposite direction than expected (e.g. systems without national curriculum standards tend to have higher average achievement scores).

Stevenson and Baker's (1991) analysis of data from 15 educational systems that participated in the Second International Mathematics Study (SIMS) eighth-grade population provides some empirical support for the argument that centralized control of the curriculum encourages more uniformity of instruction. Stevenson and Baker show that compared to educational systems in which decisions concerning the curriculum are made at the local or provincial level, systems with centralized control over the curriculum tend to have less variation in the quantity of math instruction provided across classrooms, fewer teachers who teach little of the curriculum, and more agreement among teachers in the math topics they cover in class. Additionally, the authors' analyses indicate that local factors, such as teaching experience and the heterogeneity of student ability in the classroom, exert less influence on the amount of instruction teachers provide in educational systems with centralized control over the curriculum.[12]

However, there is little reason to believe that national curriculum standards *on their own* promote uniformity of instruction. This is for several reasons. Perhaps most importantly, having national curriculum standards does not prevent nations from differentiating the curriculum within grades. This is the point Westbury and Hsu (1996) make in their critique of Stevenson and

---

survey response box: "The Organization for Educational Research and Development in Ministry of Education is the highest level of decision making. The program is mandatory throughout the country." The response suggests that they do have some form of official standards.

[12] Stevenson and Baker (1991) do not examine the relationship between uniformity of instruction and student achievement, so their work does not aid our understanding of that relationship.

Baker’s research. Westbury and Hsu contend that part of the reason why Stevenson and Baker find a relationship between central control over the curriculum and less variation in instruction across classrooms is because the centralized systems in their study tend to offer only one math course for eighth graders, while two of the five decentralized systems in their analyses offer multiple math courses at the eighth grade level. This type of curriculum differentiation should lead to greater variability of math instruction across eighth-grade classrooms—regardless of whether or not curricular decisions are made at the national level.

Focusing on the uniformity of instruction across classrooms, as opposed to indirect measures of educational inputs, such as national curriculum standards or curriculum differentiation, might bring us closer to understanding why the East Asian-U.S. achievement gap in math persists. If uniformity of instruction across classrooms is: 1) greater in East Asian nations than in the U.S., as Stevenson and Stigler (1992) assert, and 2) associated with higher average achievement at the system level, then the relative variability of math instruction across classrooms in the U.S. could be contributing to the East Asian-U.S. achievement gap. Fortunately, TIMSS gathered information from teachers on the number of math topics included on its assessment that students had been taught, which affords the opportunity to empirically address this research question.

The purpose of this chapter is to empirically assess Stevenson and Stigler’s (1992) assertion that the uniformity of math instruction across elementary-school classrooms is greater in East Asian educational systems than in the U.S. and to examine, at the educational system level, the relationship between uniformity of instruction and achievement using data from all educational systems that participated in the 2003 or 2007 TIMSS fourth-grade population. If East Asian educational systems have achieved significant uniformity of instruction across same-grade classrooms compared to the U.S., and uniformity of instruction relates to higher average achievement, then the institutional arrangements that allow for less uniformity of instruction across classrooms in the U.S. might be partially responsible for the East Asian-U.S. achievement gap.

LITERATURE REVIEW

Based on their observations of multiple classrooms in Japan, Taiwan and the U.S., Stevenson and Stigler (1992) claim that compared to the U.S., instruction is more uniform across classrooms in East Asian educational systems. But there is little empirical evidence to support or refute this hypothesis. Stevenson and Baker's (1991) study is one of the few to compare the uniformity of instruction across educational systems. Of the educational systems under investigation in this dissertation—Hong Kong, Japan, Singapore, Taiwan and the U.S.—only the U.S. and Japan are included in Stevenson and Baker's analyses. Nonetheless, Stevenson and Baker's results provide some support for Stevenson and Stigler's assertion. Stevenson and Baker show that the amount of variation (measured in standard deviations) in the number of math topics taught across classrooms is twice as large in the U.S. (20.5) as in Japan (10.0). Compared to the other educational systems in their study, the amount of variation in the quantity of math instruction provided in Japan is quite small (it ranks as the second-lowest among the 15 educational systems), while the amount of variation in the U.S. puts it in 11th place.[13] By using data from two rounds of the TIMSS fourth-grade population, this chapter builds on Stevenson and Baker's

[13] Stevenson and Baker (1991) also measure the minimum number of topics that teachers taught and the level of agreement among teachers regarding the curriculum but they do not provide descriptive statistics of these variables. For this reason, I am unable to compare results for Japan and the U.S. on these other measures of the uniformity of instruction.

work to expand the number of educational systems included in the analyses and to examine the relationship between uniformity of instruction and math achievement.

Stevenson and Stigler (1992) suggest that the lack of uniform math instruction across classrooms within the U.S. has a devastating impact on student outcomes. They explain (1992: 137), “Enormous diversity in what is taught in the nation’s schools and the fact that not all children have access to a basic core of knowledge and skills means that large numbers of young Americans cannot compete for future employment or participate fully as citizens.” Based on Stevenson and Stigler’s logic, one would expect average achievement scores to increase along with uniformity of instruction, as the latter diminishes the odds that large numbers of students do not receive instruction on basic topics and skills.

International comparative research rarely examines the relationship between uniformity of instruction and achievement. International comparative educational research on the relationship between curriculum differentiation and achievement dispersion perhaps comes closest to addressing this issue. According to this research, achievement dispersion, or inequality in achievement, should be higher in educational systems that provide different groups of students with different curriculum. This argument is based on studies of ability grouping, which show that students in higher ability groups are given more learning opportunities than students in lower ability groups, and that the difference in learning opportunities contributes to the difference in average achievement across groups (Barr and Dreeben 1983; Carbonaro 2005; Gamoran 1986, 1987). The expectation is that curriculum differentiation leads to greater dispersion of achievement.

In their review of the international comparative research on curriculum differentiation and achievement dispersion, Van de Werfhorst and Mijs (2010) report contradictory findings from several studies. The mixed results could have to do with the vague nature of the variables used to measure curriculum differentiation, such as the number of school tracks and the age of selection into tracks, which do not entirely capture the amount of curriculum differentiation that occurs within nations. For example, the amount of curriculum differentiation that occurs in the U.S. appears relatively benign compared to Japan and Singapore when studies rely solely on the age of selection into different tracks (which is 18 in the U.S., 14 in Japan and 12 in Singapore) to measure curriculum differentiation.

This chapter contributes to international comparative research by measuring the degree of inequality of educational opportunities directly, as the degree of uniformity of instruction across same-grade classrooms, and by examining the relationships between uniformity of instruction, achievement and achievement dispersion. If the hypothesis developed from studies of ability grouping is accurate, then I should find a negative relationship between achievement dispersion and uniformity of instruction across classrooms (e.g. as uniformity increases, achievement dispersion should decrease).

UNIFORMITY OF INSTRUCTION IN THE U.S.

If educational systems that have greater uniformity of instruction across classrooms tend to have higher average achievement scores, then studying the relative uniformity of math instruction provided across classrooms within states in the U.S. might help to explain why, as I demonstrate in the Introduction to this dissertation, average achievement is higher in the individual states that participate in TIMSS than it is in the nationally representative sample. Recall that in the 2007 sample, fourth-grade students in Massachusetts, on average, score higher on the TIMSS math assessment than their counterparts in Japan, and that students in Minnesota are not far behind.

Students in the nationally representative sample from the U.S., on the other hand, which does not include the students from the Minnesota and Massachusetts samples, averaged a math achievement score that is nearly one-half of a standard deviation below that of Japan. Given that the authority to determine the curriculum lies with each state, and each state uses its own assessment to measure student achievement, one would expect to find more uniformity of instruction within states than across states.

The National Central for Educational Statistics (NCES) provides some evidence that suggests that math instruction varies more across than within states. The NCES maps each state's standard for proficient performance to the achievement scale for the National Assessment of Educational Progress (NAEP). This mapping procedure allows the NCES to compare the strictness of the criteria for proficiency that each state employs. The NCES has conducted this study four times since 2003 using state data on reading and math proficiency at grades 4 and 8. In each study, the NCES finds that proficiency standards vary significantly across states (NCES 2011). For example, in the most recent study the difference between the state with the lowest standards, Tennessee, and the highest, Massachusetts, was 60 points on the NAEP math assessment. To give an idea of the magnitude of this difference, the standards Tennessee employs are lower than the basic standard set by the NCES and the proficiency standard of Massachusetts is higher than the proficiency standard set by the NCES. While the rigorousness of each state's standards do not foretell the variability in the math instruction that students in each state receive, they do suggest that there might be significant variation across states regarding the math skills and knowledge that educators believe fourth-grade students should obtain. I make use of data from the three states—Indiana, Massachusetts and Minnesota—that participated in TIMSS to conduct an exploratory analysis into the variation of math instruction within states and across states in the U.S. Unfortunately, the number of states that have participated in the TIMSS is small, which precludes a substantive analysis of the relationship between uniformity of instruction and achievement within the U.S. at this time.[14]

DATA

TIMSS is unique among internationally comparable datasets for the amount of information it gathers concerning the math curriculum in each participating nation. The TIMSS assessment frameworks delineate the math topics included in the assessments at each grade level, and the Teacher Questionnaire solicits information from the math teacher(s) of each sampled classroom on the portion of the topics included in the TIMSS assessment that students had been taught. Specifically, the questionnaire lists each math topic on the TIMSS test and asks teachers to mark whether each topic was "mostly taught before this year," "mostly taught this year" or "not yet taught or just introduced." An example topic from the 2007 fourth-grade Teacher Questionnaire reads: "Model simple situations involving unknowns with expressions or number sentences." These data are similar to those used by Stevenson and Baker (1991). The main difference between SIMS, which Stevenson and Baker used, and TIMSS is the number of math topics listed on the questionnaire. The number included in SIMS is much larger (157 topics) than the number included in either round of TIMSS (42 topics in 2003 and 35 topics in 2007).

Data for this chapter come from educational systems that participated in the 2003 or 2007 round of the TIMSS fourth-grade population and are not missing substantial information

[14] Nine states participated in the 2011 round of TIMSS. When those data become available I plan to include them in my analyses in the hopes that I will be able to contribute more insight into this research question.

regarding the TIMSS math topics that students had been taught.[15] I rely on data from the fourth-grade population, which minimizes the effect of officially sanctioned curriculum differentiation (e.g. tracking) on the variability of math instruction each educational system provides as none of the systems included in this analysis offer more than one math course to fourth-grade students. Table 2.1 lists the educational systems that are included in the analyses of each survey round.

Stevenson and Baker (1991) use educational systems, as opposed to nations, as their unit of reference. This distinction allows them to include data from places like Flemish-speaking Belgium and the Ontario province of Canada, which are not nationally representative but are representative of the educational system for those populations. I follow their lead to analyze data by educational system, as opposed to nation, which affords the opportunity to evaluate data from the various U.S. states that participated in the 2003 or 2007 round of TIMSS in addition to the nationally representative data that were also collected. Lastly, I conduct the analyses for this chapter separately by survey round as measures of the uniformity of classroom instruction are influenced by the different number of math topics included in each round's assessment. For example, the minimum proportion of topics that could have been taught is smaller in the 2003 data (1/42) than in the 2007 data (1/35).

The TIMSS sampling design calls for questionnaires and assessments to be administered to one or more math classrooms from each sampled school. This is the ideal scenario for empirically assessing the variability of math instruction across classrooms within each educational system. However, there are some instances in the TIMSS data in which the quantity of instruction that students received varies within the classroom. In these cases, some students in

[15]Kazakhstan, Morocco, Russia and Yemen excluded teacher responses to some or all questions regarding the TIMSS math topics that students had been taught and were thus excluded from analysis. I also excluded from analysis educational systems in which a third or more of students are missing data on the topics they have been taught. This step led to eliminating Armenia (56.2% of students missing data), Iran (98.8%), Latvia (52.4%), Moldova (36.8%) and Tunisia (97.0%) from the 2003 sample, and Algeria (40.9%), Georgia (39.23%), Kuwait (47.07%), Morocco (49.6%), Mongolia (50.8%), Yemen (47.8%) and Dubai, UAE (49.1%).

the classroom received extra math instruction from another teacher.[16] To use the classroom as the unit of analysis, I would need to: 1) average the quantity of instruction provided within these classrooms, which would result in inaccurate measures of the quantity of instruction each student in those classrooms received, 2) eliminate the whole classroom from the sample, or 3) eliminate the group that received additional instruction from the sample. Either of the last two options would affect the amount of variability of instruction that I measure at the system level. Although the number of instances in which the quantity of instruction varies within a classroom is small,[17] I use the student as the unit of analysis to keep the system-level measure of variability of instruction as accurate as possible.

There are other reasons as well to avoid using the teacher as the unit of analysis in the TIMSS data. In some cases, classrooms were taught math by more than one teacher. If I use the teacher as the unit of analysis, then students who were taught by more than one math teacher would be included in the analyses more than once. Further, the number of topics taught by any one of the teachers linked to the classroom would not be representative of the entire quantity of math instruction that students received. To accurately measure the quantity of instruction students in these classrooms received, I combine topic responses from their math teachers. In other cases, more than one math classroom was sampled from a school, and the same math teacher was responsible for teaching both classrooms. The quantity of instruction that these teachers provide sometimes differs across classrooms. To take the teacher as the unit of analysis

[16] It is unclear why these situations occur. They could be signs of ability grouping within classrooms, or signs that some students receive remedial (or advanced) mathematics training outside the classroom from another mathematics teacher at the school.

[17] In the 2003 data, 2 classrooms from the U.S. sample, 1 classroom from the Australia sample, and 5 classrooms from the Indiana sample include a proportion of students who were taught additional mathematics topics by another math teacher. In the 2007 data, 16 classrooms in the U.S., 4 in Japan, 7 in Singapore, 1 in Australia, 4 in New Zealand, 1 in Scotland, 4 in Sweden and 5 in Massachusetts include a small proportion of students who were taught extra mathematics topics by another mathematics teacher.

would require averaging the quantity of instruction these teachers provided across classrooms or eliminating one of their classrooms from the analysis. Given that one of the central aims of schooling is to provide equal educational opportunities to all students, it seems reasonable, given the scenarios present in the data, to measure the variability across students (as opposed to classrooms) in the quantity of math instruction received.

Finally, missing topic responses are a problem in the TIMSS data. If I include partial information about the number of topics that students were taught, I risk building less accurate measures of the uniformity of instruction within each educational system. For precision, I exclude from the analyses teacher-student pairings that are missing information on the number of TIMSS topics taught. The fourth column of Table 2.1 reports for each survey round the number of students from each educational system's sample that are included in the analyses as a percentage of the total. (The percentage that is not included represent students who are missing data from their math teachers regarding the quantity of instruction they received.)

As explained in the Introduction, TIMSS is generally referred to as a two-stage stratified cluster design, which ensures a representative sample of fourth-grade students. To obtain accurate estimates of the population mean, median and variance of all variables under consideration, it is necessary to take the TIMSS sampling design into account. To account for the stratified nature of the sample, in all analyses I use the sampling weights for students, which represent the inverse of the probability of being selected to participate in TIMSS. Clustering tends to cause correlated errors, as the residuals attached to observations within clusters are no longer independent. To take the clustered nature of each educational system's sample into account, I use Stata's suite of survey data analysis tools to conduct the analyses using the jackknife replicate weight variables provided by TIMSS. The jackknife method in TIMSS fits the model 74 times, each time dropping one or more schools (e.g. the primary sampling unit) from the estimation. The variance is estimated using the replicated point estimates.

VARIABLES

Student achievement is measured by students' scores on the TIMSS math test. Achievement scores are internationally scaled to have a mean of 500 and a standard deviation of 100. I retain this scaling metric throughout the analyses of this chapter. As is typical of large-scale educational assessments, test questions are distributed across test booklets and each student completes a portion of the total assessment. TIMSS predicts students' scores on the entire assessment using Item Response Theory (IRT). To correct for the error inherent in this imputation process, TIMSS provides five plausible values of math achievement for each student. For accuracy, I conduct each analysis five times, once on each plausible value, and combine estimates using Rubin's (1987) combination method, which is explained in the Introduction.

I use measures of the mean, median and interquartile range (IQR) in student achievement for analyses of the relationship between uniformity of instruction and student achievement at the system level. Descriptive statistics of these math achievement variables for each educational system are provided in Table 2.1. Educational systems are listed from top to bottom in each survey round by mean math achievement. Descriptive statistics for the 2003 and 2007 data are provided in the first and second panel, respectively. The last row of data in each panel reports the average across educational systems in that survey round. Mean math achievement scores vary slightly from those published in the official TIMSS literature due to the fact that the analyses are carried out on a subsample of students (e.g. those not missing data on the math topics they have been taught).

Table 2.1 Descriptive Statistics of Math Achievement and Sample Size

2003 Sample					
	Math Achievement				% in
Educational systems	Mean	Median	IQR	N Students	Subsample
Singapore	593	599	105	5,796	86.9
Hong Kong	574	577	85	4,089	88.7
Taiwan	564	567	81	4,188	89.9
Japan	563	567	96	3,923	86.5
Belgium (Fl)	550	551	82	3,533	76.6
Netherlands	541	543	72	2,106	71.7
Indiana, US	534	537	86	2,044	91.6
Lithuania	534	539	98	3,421	77.4
England	532	536	120	2,412	69.4
Hungary	530	535	105	2,509	75.6
United States	521	525	103	7,477	76.9
Ontario, Canada	512	514	93	3,225	75.8
Cyprus	511	516	113	3,319	76.7
Quebec, Canada	505	508	88	3,292	76.1
Italy	503	507	109	4,282	100.0
Australia	499	504	107	3,160	80.3
New Zealand	493	499	113	3,179	73.8
Scotland	492	494	103	2,563	68.5
Slovenia	480	486	106	2,490	79.7
Norway	451	457	108	3,328	76.6
Philippines	363	356	159	3,380	73.9
Sample Average	*516*	*520*	*102*	*3,510*	*79.7*
2007 Sample					
	Math Achievement				% in
Educational systems	Mean	Median	IQR	N Students	Subsample
Hong Kong	609	611	88	3,548	94.6
Singapore	598	604	112	4,430	88.5
Taiwan	575	577	91	3,561	86.2
Massachusetts, US	572	573	91	1,570	90.3
Japan	569	573	100	4,013	90.5
Minnesota, US	557	563	105	1,639	89.1
England	542	546	114	3,684	85.4
Latvia	539	543	95	3,150	80.6
Netherlands	534	536	82	2,371	70.8
Lithuania	530	536	102	3,217	80.8
United States	528	530	102	7,121	91.0
Denmark	525	527	92	2,496	70.9
Germany	523	527	89	3,969	76.7

Table 2.1 Continued

2007 Sample, cont.					
	Math Achievement				% in
Educational systems	Mean	Median	IQR	N Students	Subsample
Quebec, Canada	520	522	93	2,783	71.6
Australia	517	519	110	3,278	80.4
Ontario, Canada	514	516	91	2,698	77.6
Hungary	508	515	125	3,077	76.0
Austria	507	511	90	3,769	77.6
Italy	506	509	104	3,434	76.8
B.C., Canada	506	507	91	2,973	72.5
Alberta, Canada	504	506	88	3,492	86.5
Sweden	502	504	87	3,383	75.6
Armenia	502	500	121	3,237	79.4
Slovenia	501	505	95	3,161	72.6
Slovak Republic	497	504	107	4,132	83.3
Scotland	496	501	107	2,683	70.1
New Zealand	493	498	119	3,744	79.2
Czech Republic	489	492	95	3,640	86.0
Norway	474	479	102	2,950	72.4
Ukraine	469	475	114	4,292	100.0
Iran	403	407	119	2,625	68.5
Colombia	361	360	120	3,468	72.2
El Salvador	330	328	126	3,294	79.1
Tunisia	319	322	157	2,809	67.9
Qatar	297	298	127	5,155	73.5
Sample Average	*498*	*501*	*104*	*3,396*	*79.8*

As shown in Table 2.1, in the 2003 sample, mean achievement ranges from a high of 593 in Singapore to a low of 363 in the Philippines. In the 2007 sample, mean achievement ranges from a high of 609 in Hong Kong to a low of 297 in Qatar. Given that mean achievement depends on the distribution of achievement scores in the population, to isolate the effects of student achievement scores that fall on either end of the distribution, I also run the analyses on the median achievement score in each educational system. According to Table 2.1, the minor differences between the mean and median achievement of educational systems do not seriously affect the ranking of educational systems by mean achievement. For example, in the 2003 data, although the mean achievement score of Indiana state (534) is equal to that of Lithuania, the median achievement score of Lithuania (539) is higher than that of Indiana (537).

The IQR measures the inequality or dispersion in achievement scores within each educational system. According to the results in Table 2.1, in the 2003 data, dispersion in math achievement is smallest in the Netherlands (72) and largest in the Philippines (159). In the 2007 data, dispersion is smallest again in the Netherlands (82) and largest in Tunisia (157). Among East Asian and U.S. educational systems, dispersion in math achievement is largest in Singapore in both survey rounds (105 in 2003, 112 in 2007). In the 2003 data, Taiwan has the smallest dispersion in achievement, at 81, while Hong Kong has the smallest dispersion of 88 in the 2007 data. Table 2.1 indicates that measures of student achievement vary between the nationally representative sample from the U.S. and the state samples. Each individual state outperformed the nationally representative sample on the TIMSS math assessment. Dispersion in achievement is considerably smaller in the Indiana sample (86) than in the nationally representative sample (103) from 2003. In 2007, dispersion in achievement was greatest in the Minnesota sample (105) followed by the nationally representative sample (102) while the Massachusetts sample has rate of dispersion (91) that is on par with Taiwan (91) and lower than Japan (100).

Curriculum variables

To gauge the average *quantity of instruction* provided by each educational system, I add up for each student the number of TIMSS math topics teachers indicate had been either “mostly taught this year” or “mostly taught before this year” and average the results across students within each educational system. Scores on this scale could range from 0 to 42 in the 2003 data and 0 to 35 in the 2007 data, although the range of scores is much smaller in both rounds. In the 2003 data, the minimum average number of topics students had been taught is 21.3 in Norway and the maximum average is 37.2 in England. In the 2007 data, the minimum average quantity of instruction provided is 16.7 in Sweden while the maximum average is 30.8 in Singapore.

I create three variables to measure the uniformity of instruction provided within each educational system. First, I use the standard deviation in the quantity of instruction each student received to assess the *variation* in the quantity of instruction provided. Scores on this scale range from 3.7 in England to 8.1 in the Philippines in the 2003 data and, in the 2007 data, from 2.9 in Singapore to 7.15 in Qatar. Lower scores on this scale signal greater uniformity of instruction. I developed the other two measures of uniformity of instruction based on Stevenson and Baker's (1991) research. The first assesses the level of *agreement* among teachers in each educational system as to what fourth-grade students should and should not be taught. To create this measure, I add up the number of topics that 10 percent or fewer or 80 percent or more fourth-grade students had been taught this year.[18] Scores on this scale range from a minimum of 0 for several educational systems in both survey rounds to a high of 6 (for the Netherlands) in the 2003 data and 10 (for Japan) in the 2007 data. Larger values on this scale signal a greater degree of agreement among math teachers regarding the fourth grade math curriculum.

I use the minimum number of topics that students could have been taught relative to the average as my final measure of the uniformity of instruction that each educational system provides.[19] To create this measure, I divide the minimum number of topics taught by the average number of topics taught in each educational system and multiply the results by 100 to get the

[18] Stevenson and Baker (1991) use the number of topics that had been taught by 90% or more, or 10% or fewer teachers within each educational system as their measure of agreement. Due to the much smaller number of topics included in the TIMSS, I use the number of topics that had been taught to 80% or more, or 10% or fewer students.
[19] Stevenson and Baker (1991) do not take into account the average quantity of instruction provided to measure the minimum number of topics taught.

minimum percentage of the average quantity of instruction that students might receive. For example, the lowest score on this measure in the 2003 data is 26.7 percent (for the Philippines), which means that students who were taught the fewest number of topics in the Philippines received instruction on only 26.7 percent of the average number of topics that fourth-grade students in the Philippines tend to be taught. The highest score on this measure in the 2003 data is 70.0 percent (for England). In the 2007 data, scores on this scale range from a low of 0 (for New Zealand) to a high of 61.9 percent (for Taiwan). Lower scores on this scale signal less uniformity of instruction.

Table 2.2 Correlation Matrix for Measures of Uniformity of Instruction

	2003		
	Variation	Minimum %	Agreement
Variation	1.00	.	.
Minimum %	-0.69***	1.00	.
Agreement	-0.45*	0.22	1.00
N = 21			
	2007		
	Variation	Minimum %	Agreement
Variation	1.00	.	.
Minimum %	-0.70***	1.00	.
Agreement	-0.57***	0.48**	1.00
N = 35			

*** p<0.001, ** p<0.01, * p<0.05

Given that the three variables are derived from the same distributions, they are correlated by design. Table 2.2 displays the correlation matrix for the three measures of uniformity of instruction. The results show that the association between variation in quantity of instruction and minimum percentage of instruction is negative, as expected, and statistically significant at the p<0.001 level in both the 2003 and 2007 data. This finding suggests that the two variables are highly related but are measuring opposite ends of the same latent construct. The relationship between agreement and variation is also negative, as expected, and statistically significant at the $p<0.05$ level in the 2003 data and at the $p<0.001$ level in the 2007 data, suggesting that these two variables are also tapping into opposite ends of the same latent construct. The relationship between agreement and minimum percentage is positive, as expected, but the weakest of the three for both survey rounds and statistically insignificant in the 2003 data (it is statistically significant in the 2007 data, at the $p<0.01$ level). The statistical insignificance of the relationship in the 2003 data could be due to the small sample size (N=21). These facts suggest that the agreement among teachers regarding the math curriculum and the minimum percentage of topics taught are related to the variation of instruction, but are capturing different dimensions of the concept of uniformity of instruction.

To examine the relationship between achievement and uniformity of instruction in greater depth, I need to combine the three measures of uniformity to form a single variable, because, as Table 2.2 indicates, the three measures of uniformity of instruction are highly correlated. To do this, I reverse score the variation in quantity of instruction variable so that higher scores signal smaller standard deviations in the number of math topics students had been taught. Then I standardize all three variables to have a mean of zero and a standard deviation of unity before combining them into a single measure of the uniformity of instruction.

RESULTS

Is the provision of instruction more uniform in East Asia than in the U.S.?

Stevenson and Stigler (1992) contend that compared to the U.S., instruction is more uniform across classrooms in East Asian educational systems. This assertion is based on their observations of multiple classrooms in Japan, Taiwan and the U.S. I empirically assess the difference in the uniformity of instruction provided in multiple East Asian educational systems and the U.S. by utilizing nationally representative data from Hong Kong, Japan, Singapore, Taiwan and the U.S., as well as data that are representative of the fourth-grade population in Indiana, Massachusetts and Minnesota. Table 2.3 displays the average quantity of instruction students in each of these educational systems received, the degree to which the quantity of instruction varied across students, the minimum percentage of topics students were taught and the extent of agreement among math teachers regarding the math curriculum. Results from the U.S. data, which includes separate datasets for Indiana, Massachusetts and Minnesota in addition to the national sample, are displayed in the first four rows of Table 2.3. The next four rows of data provide results from East Asian educational systems. The second to last row of data gives the average from the sample of eight East Asian and U.S. educational systems. The last row of data provides the sample average from all educational systems that participated in each round.

Table 2.3 Quantity and Uniformity of Instruction in East Asia and the U.S.

	Quantity of Instruction		**Variation**		**Minimum Percentage**		**Agreement**	
Educational Systems	2003	2007	2003	2007	2003	2007	2003	2007
United States	34.5	30.0	6.4	4.8	29.0	26.7	5.0	5.0
Indiana, US	32.0	.	5.8	.	65.6	.	2.0	.
Massachusetts, US	.	29.2	.	5.0	.	37.7	.	7.0
Minnesota, US	.	29.4	.	5.3	.	44.3	.	4.0
Hong Kong	30.7	26.2	6.1	3.7	45.5	57.2	3.0	6.0
Japan	22.7	21.4	5.3	4.0	39.6	46.7	6.0	11.0
Singapore	34.7	30.8	4.2	2.9	57.7	61.7	8.0	9.0
Taiwan	32.8	27.5	5.0	4.6	54.9	61.9	7.0	1.0
East Asia-U.S. Sample Average	31.2	27.8	5.5	4.3	48.7	48.0	5.2	6.1
Full Sample Average	31.1	24.5	5.5	4.9	46.6	38.7	4.3	3.9

The first two columns of Table 2.3 show the average quantity of instruction students in each educational system received in 2003 (column 1) and 2007 (column 2). This measure tells us the average size of the implemented curriculum in each educational system. The East Asian-U.S. sample average in the 2003 data is 31.2 (out of 42, or 74.4%), which is nearly identical to the average for the entire sample of 21 systems (31.1, or 74.1%). In the 2007 data, the East Asian-U.S. sample average is 27.8 (out of 35, or 79.4%), which is higher than the average for the entire sample of 35 educational systems (24.5, or 70.1%).

The results in Table 2.3 indicate that students in Singapore, on average, were taught the largest number of topics (34.7 or 82.6% in 2003, 30.8 or 88.0% in 2007) followed by students in the U.S. (34.5 or 82.1% in 2003, 30.0 or 85.6% in 2007), Minnesota (29.4 or 83.9%), Massachusetts (29.2 or 83.3%), Taiwan (32.8 or 78.0% in 2003, 27.5 or 78.5% in 2007), Indiana (32.0 or 76.2%) and Hong Kong (30.7 or 73.2% in 2003, 26.2 or 74.9% in 2007). In both rounds, students in Japan typically received instruction on significantly fewer topics (22.7 or 54.1% in 2003; 21.4 or 61.2% in 2007) than their counterparts in East Asia and the U.S. This finding is particularly interesting in light of Stevenson and Baker's (1991) work, which shows that out of all educational systems in their sample, Japan averaged the largest number of topics taught. The difference between the findings of this chapter and Stevenson and Baker's likely reflect the extensive curriculum reforms that have taken place in educational systems around the world since the 1980s. The results across both rounds of data show that, on average, all of the educational systems in East Asia and the U.S.—except for Japan—provide a greater-than-average amount of math instruction.

The next two columns of Table 2.3 tell us the degree of variation in the number of topics taught in 2003 (column 3) and 2007 (column 4). The East Asia-U.S. sample average variation is

5.5 in the 2003 data, which is equivalent to the average for the entire sample. Variation in the 2003 data is greatest in the U.S. (6.4), and is 152 percent that of Singapore (4.2), the educational system with the smallest variation in instruction. Interestingly, in the 2003 data, variation in the quantity of instruction students received is larger in Hong Kong (6.1) than in Indiana (5.8), which suggests that the variability of instruction is not always consistently lower in East Asia than in parts of the U.S. Variation is smaller than the average for the entire sample in Japan (5.3) and Taiwan (5.0), which is consistent with Stevenson and Stigler's assertion that these educational systems have achieved significant uniformity of instruction.

In the 2007 data, shown in column 4 of Table 2.3, variation in the quantity of instruction that students received is again smaller than the entire sample average in Japan (4.0) and Taiwan (4.6); smallest among all East Asian and U.S. educational systems in Singapore (2.9), and greatest for students in parts of the U.S. (5.3 in Minnesota and 5.0 in Massachusetts). Across rounds, variation declined in the nationally representative U.S. sample from the largest amount in the 2003 sample to 4.8 in 2007, which is below the entire sample average and close to that of Taiwan. A similar trend can be seen in the Hong Kong data. Variation in the 2003 data for Hong Kong was higher than the entire sample average, but in 2007 it is smaller (3.9) than the average for the entire sample. Despite the decrease in variation in the U.S. across rounds of data, variation in the quantity of instruction in 2007 is still larger in all U.S. educational systems than East Asian systems. Additionally, the results in these columns do not support the hypothesis that math instruction varies less within than across states.

Columns 5 and 6 of Table 2.3 indicate the minimum percentage of the average quantity of instruction provided in each educational system. The East Asian-U.S. sample average on this scale is 48.7 percent in the 2003 data, which is slightly higher than the average of 44.6 percent for the entire sample. Interestingly, the minimum is both highest and lowest in the U.S. in the 2003 data. In the nationally representative sample, students who received the least amount of instruction were taught only 29.0 percent of the average while students in Indiana who were taught the fewest number of topics received instruction on 65.6 percent of the average number of topics taught. Among East Asian educational systems, the minimum percentage is below the entire sample average in Japan (39.6%) and Hong Kong (45.5%), and above the entire sample average in Singapore (57.7%) and Taiwan (54.9%). In the 2007 data, the East Asia-U.S. sample average of 48.0 percent is well above the entire sample average of 38.7 percent. In this round, the minimum percentage taught is again lowest in the nationally representative sample from the U.S. (26.7%). The states that participated in this round had the second and third lowest scores, of 37.7 percent for Massachusetts and 44.3 percent for Minnesota. The minimum percentage taught in all East Asian educational systems is above the entire sample average. The minimum is highest in Taiwan (61.9%) followed by Singapore (61.7%), Hong Kong (57.2%) and Japan (46.7%). The data from 2007 conform to expectations more so than the data from 2003.

The last two columns of data of Table 2.3 tell us the number of topics that were taught mostly this year to either 80 percent or more or to 10 percent or fewer students in each educational system. In the 2003 data, the East Asia-U.S. sample average is 5.2, which is larger than the average of 4.3 for the entire sample. On this measure of agreement, Indiana fares the worst, with a score of 2, followed by Hong Kong (3). The remaining educational systems scored above the average for the entire sample. The level of agreement is highest in Singapore (8) followed by Taiwan (7), Japan (6) and the U.S. (5). In the 2007 data, the average of 6.1 for the East Asian-U.S. sample is again higher than the average of 3.9 for the entire sample. Interestingly, East Asian educational systems have the highest and lowest scores on this scale.

Japan's score of 11 is highest among all educational systems that participated and suggests a high level of agreement among teachers regarding the material that students should and should not be taught. The number of topics on which teachers agree is smallest in Taiwan (1). Minnesota's score of 4 is roughly equivalent to the entire sample average. After Japan, the educational system with the largest score is Singapore (9), followed by Massachusetts (7), Hong Kong (6) and the U.S. (5).

Taken together, the results of Table 2.3 support Stevenson and Stigler's (1992) assertion that East Asian educational systems have achieved significant uniformity of instruction. The results indicate that the quantity of instruction provided within the U.S. varies considerably more than it does within any East Asian educational system. Further, the minimum percentage of instruction that students in the U.S. might receive is much smaller than the minimum percentage that students in East Asian educational systems might receive. Additionally, there tends to be more agreement among math teachers regarding the curriculum in East Asia than in the U.S. The notable exception is Taiwan, where agreement is low in the 2007 data. In short, the nationally representative sample from the U.S. ranks at or near the bottom of every measure of uniformity of instruction in each survey round. Among East Asian educational systems, math instruction appears to be most uniformly distributed in Singapore, as Singapore had the first- or second-best score on each measure of uniformity of instruction in both rounds of data.

The picture regarding the uniformity of math instruction within states as opposed to across states is less clear. Each state that participated in TIMSS ensured that its students received a higher minimum percentage of instruction relative to the average than the U.S. Beyond that, however, the results are not conclusive. Given the small sample size, it is likely that increasing the number of states in the analyses would shed more light on this research question.

Is uniformity of instruction related to achievement?

Results from the previous section suggest that the uniformity of math instruction is greater in East Asian educational systems than in the U.S. The next question is whether uniformity of math instruction is related to higher average math achievement. Evidence of a positive relationship between uniformity of instruction and average achievement would be consistent with the argument that greater variability of math instruction in the U.S. compared to East Asia is contributing to the East Asian-U.S. achievement gap. To empirically examine this research question, I first calculate Pearson product-moment correlations between measures of uniformity of instruction and achievement at the educational system level in each survey round.[20] To test the statistical significance of these relationships, I calculate the t-test probability statistics for the null hypothesis that each individual correlation equals zero. Table 2.4 reports the results from these analyses. The first two columns of data in Table 2.4 give the Pearson product-moment, *r* , between mean achievement and each of the three measures of uniformity in instruction in the 2003 round (column 1) and 2007 round (column 2). The next two columns show results for the relationships between measures of uniformity and median student achievement. The final two columns of data report the *r* for each relationship between achievement dispersion and measures of uniformity.

The results displayed in Table 2.4 indicate a strong, negative and statistically significant relationship between variation in the quantity of math instruction provided and average math achievement (measured using either the mean or the median) in both survey rounds. This finding suggests that average student achievement tends to be higher in educational systems with less

[20] I ran the analysis using data from each U.S. state as well as from the national sample and then again on a sample that excludes the U.S. states. Results from each analysis are similar (e.g. relationships are in the same direction and have the same statistical significance). Results reported here are from the analysis that includes each U.S. state.

variation in the quantity of instruction provided, and this relationship is not sensitive to the survey year or sample size. There is some evidence, too, that average achievement is higher in educational systems where students receive a higher minimum percentage of instruction. In both survey rounds, the minimum percentage of instruction that students might receive is positively and statistically significantly related to average student achievement (measured as either the mean or the median).

Table 2.4 System-level Correlates between Achievement and Uniformity

	Student Achievement					
	Mean		Median		Dispersion	
Variables	2003	2007	2003	2007	2003	2007
Variation	-0.65**	-0.65***	-0.66***	-0.64***	0.40^	0.34*
Minimum Percent	0.45*	0.45**	0.47*	0.44**	-0.22	-0.26
Agreement	0.26	0.23	0.25	0.21	-0.07	0.12
N	21	35	21	35	21	35

*** p<0.001, ** p<0.01, * p<0.05, ^ p<0.10

Table 2.4 shows that the relationships between indicators of average student achievement and agreement (measured as the number of topics that were taught to 80% or more or 10% or fewer students) are in the expected direction. The relationships are positive, suggesting that average achievement scores increase when teachers are in agreement regarding the math curriculum. However, the relationships are not statistically significant, which could be due to the relatively small sample size in each survey round (21 educational systems in the 2003 data and 35 in the 2007 data). In these data, it seems safe to say that the relationships between average achievement and measures of variation and minimum percentage are stronger than the relationship between agreement and achievement.

The relationships between achievement dispersion and measures of uniformity of instruction are all in the expected direction. The results of Table 2.4 suggest that dispersion in achievement increases with the variability of instruction and decreases with the minimum percentage of instruction and with agreement among teachers. However, the only relationship that is statistically significant is between variation and achievement dispersion. In both rounds, the relationship is statistically significant (at the $p<0.10$ level in 2003 and at the $p<0.05$ level in 2007). Table 2.4 also sheds light on a question raised by internationally comparative research concerning the relationship between curriculum differentiation and inequality of achievement. The results indicate that inequality in achievement tends to be greater in educational systems with more variation in the quantity of instruction provided.

Overall, the results of Table 2.4 show that average achievement scores are higher in educational systems with less variation in the quantity of instruction provided, and in systems with higher minimum percentages of instruction. These findings provide some support for the argument that the East Asian-U.S. achievement gap persists in part because there is less uniformity of math instruction in the U.S. compared to East Asia.

To gain greater insight into the relationship between uniformity of instruction and achievement, I conduct OLS regression analyses of average achievement and achievement dispersion on the combined measure of uniformity of instruction. Table 2.5 reports the results from these analyses. The table is divided into three panels, one each for regressions on the mean, median and achievement dispersion. Results from the 2003 sample are in the first column of data, results from the 2007 sample are in the second. The last row of each panel provides the adjusted R-squared for each model. Standard errors are given below the coefficient for uniformity.

Table 2.5 OLS Estimates of Uniformity on Achievement

Variables	2003	2007
	Mean	
Uniformity	11.87***	14.45***
	(1.99)	(2.03)
Adj. R-squared	0.286	0.248
	Median	
Uniformity	12.07***	14.59***
	(1.98)	(2.09)
Adj. R-squared	0.297	0.232
	Dispersion	
Uniformity	-2.21	-1.19
	(1.31)	(1.03)
Adj. R-squared	0.037	0.008

***p<0.001; Standard errors are in parentheses.

Results displayed in Table 2.5 suggest that educational systems with more uniformity of math instruction have higher average math achievement scores than educational systems with greater variability of math instruction. In both the 2003 and 2007 round of TIMSS, uniformity of instruction is positively and statistically significantly (at the $p<0.001$ level) related to average achievement, measured as either the mean or the median. Uniformity of instruction explains around 30 percent of the variance in mean and median student achievement in the 2003 round and between 23 and 25 percent in the 2007 sample.

Results of the regressions of achievement dispersion on uniformity of instruction show a relationship in the expected direction. However, the coefficients are not statistically significant, and the variable explains little of the variance in achievement dispersion. Given the findings from Table 2.4, which show that achievement dispersion is only statistically significantly related to variation in the quantity of instruction provided, these last OLS regression results are not surprising. Achievement dispersion appears to be mildly related to the variability in the quantity of instruction provided but not to the minimum percentage of instruction that students might receive nor to the variable indicating the level of agreement among teachers. These results are not surprising, given the mixed results of previous research. They suggest, perhaps, that the amount of achievement dispersion in an educational system depends not only on the degree of curriculum differentiation within the system but also on the extent to which students within the system differ by ability and effort. In other words, if an educational system could achieve significant uniformity in both the quantity and quality of instruction it provides, the system would still have some degree of achievement dispersion due to differences across students in ability and effort (including parental support and involvement).

CONCLUSION

The principal aim of this chapter is to empirically assess the possibility that average math achievement is higher in East Asia than in the U.S. because math instruction is more uniformly distributed in East Asian educational systems. Using representative data from educational systems that participated in the 2003 and 2007 TIMSS fourth-grade population, I provide empirical evidence consist with this hypothesis. I show that variation in the quantity of math instruction students receive is smaller in all East Asian educational systems than in the U.S.; that the minimum percentage of math topics students might be taught is higher in East Asian educational systems than in the U.S., and that there tends to be more agreement on the fourth grade math curriculum among teachers in East Asian educational systems than in the U.S. All of these findings support Stevenson and Stigler's assertion that compared to the U.S., East Asian educational systems have achieved significant uniformity of instruction across classrooms.

This chapter suggests that uniformity of instruction is related to higher average achievement because uniformity ensures that all students receive at least a basic education. The analyses of this chapter provide some support for this notion. They show strong relationships at the system level between average achievement and the amount of variation in the quantity of instruction provided as well as between achievement and the minimum percentage of instruction relative to the average that students might receive. These relationships were statistically significant across rounds. Educational systems with more variation in the quantity of math provided tend to have lower average math achievement scores. Additionally, educational systems that ensure all students receive at least a basic education tend to have higher average achievement scores than those in which students might be taught a small fraction of the topics that students in that system typically learn.

The evidence provided in this chapter is consistent with the argument that the institutional arrangements in the U.S. that allow for greater variability of math instruction might be partially responsible for the East Asian-U.S. achievement gap. However, the results are far from conclusive. It might be the case that, among educational systems that participate in TIMSS, uniformity of instruction goes hand-in-hand with more finely-tuned curricular guidelines, textbooks, teacher-training manuals or other variables that influence the quality of the curriculum but are not included in the analyses of this chapter. Thus, before policy recommendations can be made, work should be done to learn more about the similarities and differences in the math curriculum offered in educational systems with uniformity of instruction compared to those without.

While this chapter has shown that the uniformity of instruction is greater in East Asia than in the U.S., it does not examine the institutional arrangements that might be promoting greater uniformity of instruction in East Asia and other educational systems. Future research might try to empirically assess the policy measures and practices that relate to uniformity of instruction. This chapter has paved the way for this type of research by showing how the TIMSS data can be used to measure the uniformity of instruction provided by educational systems.

There are some factors that might influence the relationship between uniformity of instruction and student achievement that this chapter does not address. For example, the quality of the curriculum could vary among educational systems with the same degree of uniformity of instruction, which might influence the relationship between uniformity of instruction and student achievement. Further, it is possible for educational systems to achieve a high degree of uniformity of instruction and not provide equal educational opportunities to all students. This happens when all students receive the same quantity of instruction, but the *quality* of instruction differs. For educational systems to provide equal learning opportunities to all students, variation in both the quantity and quality of instruction should be small. The next chapter examines this research question directly, by showing how teacher effectiveness in the fourth grade varies within and across nations.

Chapter 3

Teachers

OF THE ACTORS INVOLVED in the process of student learning, other than students math teachers primarily are responsible for student achievement on a math test. The centrality of teachers to student learning, coupled with the fact that the least proficient 5 percent of students in East Asian nations tend to perform on par with the average U.S. student on the TIMSS math assessments, provides reason to believe that math teachers in East Asian nations do a better job than their U.S. counterparts of promoting achievement. Currently, there is limited empirical evidence to support or repudiate this hypothesis. The purpose of this chapter is to empirically address the claim that math achievement is higher in East Asian nations in part because math teachers there are more effective than their U.S. counterparts.

International comparative research provides compelling reasons to believe that math teachers in East Asian nations are more effective than those in the U.S. Observational studies show that math teachers in East Asia tend to introduce conceptually challenging problems, and work to develop their students' knowledge of underlying procedures (Stevenson and Stigler 1992; Stigler and Hiebert 1999; Ma 1999; Perry 2000; Leung 2005). By contrast, teachers in the

U.S. tend to present problems that are less intellectually stimulating (Leung 2005), and their lessons tend to involve mainly memorization of mathematical rules and terms, and drill (Stigler and Hiebert 1999). These studies argue that even when teachers are teaching the same material, cross-national variations in instructional strategies ensure that students in East Asia develop a richer understanding of the subject matter. In other words, East Asian teachers are more effective due to the high quality of their instruction.

This scholarship also highlights structural features of East Asian educational systems that act to produce a more effective teaching corps. These features include lighter instructional workloads (Stevenson and Stigler 1992), continuing professional support and development (Stigler and Hiebert 1999; Akiba and LeTendre 2009), formally scheduled time to plan lessons (Stevenson and Stigler 1992), better teacher compensation (Stevenson and Stigler 1992), and more rigorous teacher training and selection programs (Wang et al. 2003). Wang et al.'s (2003) exploratory analysis of national differences in policies related to teacher training and development demonstrates that, compared to other nations that scored as well or better than the U.S. on the 1999 TIMSS achievement tests, the U.S. employs a less robust teacher education and development process.[21] Hong Kong, Japan, and Singapore (Taiwan was not included in their analysis), for example, formally monitor teachers' performance during the induction period, and/or their professional development. Additionally, in Hong Kong and Japan, tenure is not automatically granted. The U.S., on the other hand, lacks such filters (Wang et al. 2003).

[21] Although the U.S. does not have a national teacher certification program, 42 states participate in a reciprocal teacher certification agreement through the National Association of State Directors of Teacher Education and Certification (NASDTEC), which makes it easier for teachers certified in one state to become certified in another. The states that do not participate in this agreement are concentrated in the Midwest (Iowa, Minnesota, Missouri, Nebraska, North Dakota, South Dakota and Wisconsin) and Alaska. Features such as this reciprocity program and influential accreditation organizations "result in teacher education and certification systems that are more alike than different" (Wang et al 2003: 4).

The main thrust of this research is that teachers in East Asian nations are more effective than their U.S. counterparts because East Asian educational systems select more promising candidates and/or provide better training and support structures, which in turn encourage high-quality instruction. According to this research, we should find that the overall level of teacher effectiveness is high in East Asia and that, within each East Asian nations, teachers vary less in their ability to promote achievement.

Curriculum specialists, on the other hand, contend that U.S. teachers are equally as effective as their counterparts in East Asian nations. According to Westbury (1992), the reason why Japanese students tend to outperform U.S. students on international assessments in math is because the mandated math coursework for all eighth and twelfth graders in Japan is more advanced than that completed by the majority of U.S. students. To empirically support this argument, Westbury (1992) uses data from the Second International Mathematics Study (SIMS) to show that there is no real difference in the achievement scores of U.S. students who received a similarly amount of math instruction to their Japanese counterparts. This finding has important implications for teachers, as Westbury (1992: 21) explains: "[W]hen the curricula of the two countries are similar, there is essentially no difference in the performance of U.S. and Japanese students, and, by implication, of teachers." Westbury contends that the curriculum structure in the U.S. is to blame for the U.S.'s relatively low achievement scores; if all U.S. students were given the challenging math coursework typically reserved for college- and enrichment-track students, then average achievement would be higher and it would be clear that U.S. teachers are equally as effective as their East Asian counterparts at promoting achievement.

Westbury (1992) relies on the maxim that the more students are taught, the more they will learn and the better they will score on international assessments. According to a National

Center for Educational Statistics (NCES) report prepared by Medrich and Griffith (1992), findings from the first four international assessments in math and science provide evidence in support of this idea. Sørensen and Morgan (2000: 149-50) make a similar argument: “Schools that cover extensive material in their instruction provide many opportunities for learning and will create more growth in achievement.” The authors allow for the possibility that individual teachers might influence how much students learn, but only in so far as teacher effort partially determines the quantity of material covered. (In this formulation, formal curricular guidelines act as the primary determinant of the quantity of material taught in the classroom.) Once the quantity of instruction is taken into account, variations in student effort and ability explain variations in student achievement. The implication here is that student achievement depends far more on the *quantity* of instruction than on the *quality* of instruction.

The idea that the quantity of instruction largely determines student learning and by implication, teacher effectiveness, presents an interesting foil to U.S.-based studies of teacher effects (for example: Nye, Konstantopoulos and Hedges 2004; Rivkin, Hanushek and Kain 2005; Kane, Rockhoff and Staiger 2006). These studies tend to ignore curricular differences across classrooms. Instead, teacher effectiveness is measured as the amount of variance in (residualized) gains in student achievement that is due to differences between classrooms within schools. According to results from these studies, U.S. teachers vary tremendously in their ability to promote achievement. To put the magnitude of the difference between effective and less effective teachers in the U.S. into perspective, Palardy and Rumberger (2008) offer the following comparisons: being assigned to an effective teacher has five times greater the effect of family socioeconomic background and two-and-a-half-times greater the effect of reducing class size from 25 to 15 students on students’ math achievement.

Westbury's (1992) research and Sørensen and Morgan's (2000) theory raise the possibility that U.S. teachers vary less in their ability to promote achievement than teacher effects research leads us to believe. This is because some of the variation in teacher effectiveness might be due to differences across classrooms within schools in the amount of material covered. Since the datasets used to compute teacher effects generally do not contain information on the test material that teachers covered in class (Palardy and Rumberger 2009; Sørensen and Morgan 2000), the extent to which differences in quantity of instruction influence the amount of variation in teacher effectiveness is an empirical question.

Although Westbury (1992) and Sørensen and Morgan (2000) make a good case that student learning increases with quantity of instruction, this hypothesis is hotly contested among educators. There are generally considered to be two competing philosophies regarding optimal curriculum coverage. On the one hand, there is the "breadth" or "full coverage" philosophy that Sørensen and Morgan's (2000) theory illustrates. On the other is the "deep coverage" or "depth" philosophy, whose proponents contend that some topics within a discipline are more important than others and focusing solely on those topics is more beneficial to student achievement than covering as many topics as possible. This debate raises the possibility that there might not be a systematic relationship between quantity of instruction and student achievement, and in turn, between the amount of material teachers cover and their ability to promote achievement. Which topics are taught, in which sequence and in how much depth could be very relevant to understanding variation in teacher effectiveness both cross-nationally and—given decentralized nature of curriculum in the U.S.—within the U.S. as well.

The relationship between quantity of instruction and student achievement becomes less clear when we examine aggregate data from Japan and Singapore. For fourth-grade math, Singapore might be seen as a prime example of the “breadth” philosophy while Japan stands as an archetype of the “depth” camp. As we saw in the curriculum chapter, the average student in Singapore receives instruction on about 85 percent of the TIMSS test topics while the average student in Japan receives instruction on roughly 58 percent. In both nations, average math achievement is roughly equivalent. In the aggregate, therefore, both nations stand as exemplars for each philosophy. Because there is variation among classrooms within Japan and Singapore in the amount of material covered, however, the relationship between student achievement and quantity of instruction *within* each nation is uncertain.

This chapter capitalizes on the variation in the quantity of instruction that exists within each nation to assess whether U.S. teachers who cover the same average percentage of test topics as their Japanese and Singaporean counterparts are equally effective at increasing achievement. If student achievement is strongly dependent upon the *quality* of instruction, as Stigler and Hiebert (1999) and others contend, then even after controlling for quantity of instruction I should find that math teachers in East Asian nations are more effective than their U.S. counterparts.

The literature suggests not only that math teachers in East Asia are more effective than their U.S. counterparts, but also that there is less variation among teachers within each East Asian nation in their ability to promote achievement. To empirically address this question, I utilize multilevel “value-added” models to assess the proportion of variance in student achievement that is due to classrooms within schools, separately for the U.S. and each East Asian nation. Before describing these analyses and results, I briefly review the ways that teacher effectiveness is typically measured in international comparative studies and the method preferred in U.S. research.

HOW IS TEACHER EFFECTIVENESS MEASURED?

International comparative research tends to rely on teacher background characteristics to assess teacher quality. Montt (2011), for example, uses the proportion of teachers with ISCED 5a degrees as measures of teacher quality at the school and national level. Research conducted by Akiba and colleagues (2007; Akiba and LeTendre 2009) also defines a high-quality teacher *a priori* as a teacher with full teaching certification, more than three years of experience and certified subject-matter knowledge (e.g. whether or not teachers majored in math or math education). In other words, international comparative research tends to equate high-quality teachers with highly *qualified* teachers.

The dominant viewpoint among U.S.-based educational researchers is that easily measureable teacher characteristics, such as teacher qualifications, do not readily distinguish effective teachers from less effective ones (Hanushek 2011; Hanushek and Rivkin 2010; Gordan, Kane and Staiger 2006; Kane, Rockhoff and Staiger 2005; OECD 2005; Palardy and Rumberger 2008). Several studies that analyze the relationship between student achievement and teacher qualifications do not find a significant relationship (for example: Nye et al. 2004; Palardy and Rumberger 2008, Rivkin et al. 2005). When research does find a statistically significant relationship between measures of teacher qualifications and student achievement, it is not particularly strong. For example, Darling-Hammond and colleagues (2005) found a statistically significant difference in achievement gains between the group of students taught by certified teachers and the group taught by uncertified teachers. However, the difference is small—one-half a percentile point—and is, as Gordon et al. (2006: 9) point out, “dwarfed by the differences within groups.”

Research also indicates that teaching experience is not significantly related to student achievement beyond the initial years of teaching (Gordon et al. 2006; Nye et al. 2004; Rivkin et al. 2005). Beginning teachers tend to be less effective at promoting achievement than teachers with three or more years of experience, but there is still a considerable range of effectiveness among experienced teachers (Gordon et al. 2006). Warns the OECD (2005: 27), "in light of the lack of strong evidence linking teacher credentials such as qualifications and experience to student results, alternative indicators of teacher quality are crucial."

Multilevel, "value-added" models of teacher effects circumvent the issue of predicting teacher effectiveness directly using measured teacher characteristics by computing teacher effectiveness indirectly, as the amount of variance in (residualized) student achievement gains that is not explained by student characteristics. The more sophisticated of these "value-added" models recognizes the differences among schools in student achievement and uses three-level hierarchical models in which students are nested in classrooms within schools. These models allow researchers to assess the relative explanatory power of schools, classrooms within schools (or teachers) and students on student achievement gains. There appear to be few studies that examine cross-national differences in teacher effectiveness using three-level hierarchical linear models. This chapter represents a pioneering effort to examine cross-national differences in the proportion of variance in student achievement that can be attributed to classrooms within schools.

In an ideal setting, value-added estimates of individual teachers' effectiveness would use information on student achievement on a pre- and post-test, and the contents of both tests would correspond to the curriculum that teachers cover in class. This way, achievement differences between students' pre-test and post-test scores would tell us how much students learned from what their teachers taught. Most large-scale educational data sets, however, rely on “general use” assessments, such as the Stanford Achievement Test (SAT), which were not designed to assess students’ abilities on a specific classroom curriculum. They were instead designed to cover a broad range of concepts that were deemed appropriate for each grade level.

Two problems arise from using student outcomes on general assessments to measure teacher effectiveness. The first has to do with the design of the test. General-use test publishers produce a different test for each grade level. The content of each grade’s test differs from the next, which means that the pre- and post-test achievement scores that researchers use to calculate achievement gains actually measure students’ performance on tests that cover different material. If the concepts differ across tests, or if the weights given to each concept differ across grade levels, then measures of teachers’ valued-added will be biased (Lockwood et al. 2007; Martineau 2006). Further, student achievement scores on these tests are vertically scaled from Kindergarten to the twelfth grade so that researchers can ascertain student achievement gains over the long term. A fundamental assumption of vertical scaling is that the tests being scaled measure the same concepts. The differences between the mathematical concepts on the Kindergarten test and those on the twelfth-grade test are great enough, as Martineau (2006) points out, to make even psychometricians seriously question the accuracy of such practices. This is important because inaccuracies in the scaling procedure can lead to inaccuracies in the amount of variance in achievement gains that is attributed to teachers.

The second concern researchers have expressed over the use of general assessments to measure teacher effects stems from a lack of information on the correspondence between the test curriculum and the classroom curriculum. Most educational data sets do not include information from teachers on the specific test topics they taught. Since the correspondence between the class curriculum and the test curriculum is unknown, the appropriateness of the test for measuring teachers' impact on student achievement is uncertain (Lockwood et al. 2007; Martineau 2006). The test could be measuring students' knowledge of topics they were taught in previous years, for example, or encompass only a narrow range of the topics teachers covered. Since the aim of value-added research is to adjudicate teacher ability to produce achievement, the instrument used to measure achievement needs to accurately reflect what the teacher taught. Without knowing the correspondence between the assessment curriculum and the classroom curriculum, effective teachers might be misidentified as ineffective (or vice versa). Although these matters deserve attention, as Hanushek and Rivkin (2010) point out, few teacher-effects studies question the appropriateness of the achievement tests to measure teachers' effectiveness.

While the TIMSS assessments could be considered "general use," the datasets differ from others because they provide information from teachers regarding the test topics their students have been taught. This information offers insight into the correspondence between the test curriculum and the classroom curriculum. As such, I am able to compare teachers across nations who taught similar percentages of test topics, which I do in the next section. These data also afford the opportunity to contribute to the debate on the relationship between quantity of instruction and student achievement. Lastly, information on the correspondence between the test curriculum and the classroom curriculum allows me to address the question of the appropriateness of the TIMSS math test for gauging teacher effectiveness. By running multilevel models of teacher effects on the full sample, and again on the subsample of students for whom the test appears to be a good fit with the math curriculum their teachers covered, I demonstrate the extent to which the correspondence between the test curriculum and classroom curriculum impact estimates of teacher effects.

VARIABLES

The outcome variable is student's overall achievement on the TIMSS math assessment. As I explained in the Introduction to this dissertation, achievement scores are internationally scaled to have a mean of 500 and a standard deviation of 100. I retain this scaling metric throughout the analyses in this chapter. Students' scores were imputed using Item Response Theory (IRT). To correct for the error inherent in the imputation process, TIMSS provides five plausible values for each student. For accuracy, I conduct each analysis five times, once on each plausible value, and combine estimates using Rubin's (1987) combination method, which I provide in the Introduction to this dissertation.

Neither round of TIMSS is longitudinal, so I do not have a measure of student achievement in a prior year to use to measure gains in achievement. (This information is not available in any current, international educational data set.) I have not yet found an international comparative study of student achievement that controls for student ability to estimate relationships between student achievement and student, school, and national factors. I therefore propose to use students' self-reports of how well they usually do in math to measure their prior ability.[22] By standardizing this variable on the national mean, I control for potential cross-national variations in the ways that student represent their past achievement in math. This variable has a mean of zero and standard deviation of one. Since these are not direct measures of pre-test achievement, I follow Nye et al.'s (2004) lead and acknowledge that I am examining teacher effects on achievement *status* as opposed to achievement gains.

[22] I assume that this measure is about as accurate as students' self-reports of their grades. Meta-analyses conducted on U.S. data to assess the reliability of students self-reports of grades vary substantially in their conclusions (for a review see Kuncel et al 2005.) According to Kuncel et al (2005), students' self-reports on grades are highly accurate for high ability students and less accurate for low ability students. It is unclear whether there is a pattern to self-reported measures of ability in East Asian nations.

To assess the quantity of instruction students received, I use data from the Teacher Questionnaire. The questionnaire lists, by content domain, the main topics that appear on the TIMSS assessment and asks teachers to mark whether a test topic was “mostly taught before this year,” “mostly taught this year” or “not yet taught or just introduced.” On the 2003 test, there are five math content domains: Number, Patterns and Relationships, Measurement, Geometry and Data. The 2007 round divides topics into three content domains: Number, Geometry, Shapes and Measures, and Data Display. An example test topic from the Geometry domain of the 2003 questionnaire reads: “Congruent triangle (i.e., same shape and size).” The 2003 round includes 42 main topics; 35 topics are included in the 2007 round.

It seems unlikely that each topic was given equal weight on the assessment. For this reason, I use the distribution of score points across content domains to measure the percentage of points each topic is worth. Specifically, I assign a weight to each topic based on the percentage of assessment points given to its content domain and the number of topics in the domain. This information is displayed in Table 3.1. For example, the 12 topics within the Number domain were worth a combined 68 points, which represents 40.24 percent (68/169) of the total. To calculate the percentage of points each topic in the Number domain is worth, I divide 40.24 percent by 12 to get 3.35. The value of each topic thus varies by round and content domain.

I calculate the quantity of instruction students received by first adding up the number of topics within each content domain that students were taught in the test year and in the previous year (e.g. in the third or fourth grade). Then I multiply the sum for each domain by its corresponding weight. The results indicate the “assessment-weighted” percentage of topics students have been taught. For example, according to this scale participants in the 2007 round who were taught 10 Number topics, 10 Geometric shapes topics and 5 Data Display topics received instruction on ((10x2.686) +(10x3.078) +(5x3.021)) 72.7 percent of the test topics. For the multilevel models, I center this variable on 75 percent, which is roughly the average percentage of topics taught across all countries. I chose a common mean to aid cross-national comparisons of point estimates.

Table 3.1 Information Used to Assign Assessment-Specific Values to each Topic

Round	Content Domain	Number of topics	Point distribution	% of total points	Value for each topic
2003	Numbers	12	68	40.2%	3.353
	Patterns	6	25	14.8%	2.465
	Measurement	6	33	19.5%	3.254
	Geometry	11	25	14.8%	1.345
	Data	7	18	10.7%	1.522
	Total:	**42**	**169**		
2007	Number	19	98	51.0%	2.686
	Geometric Shapes	11	65	33.9%	3.078
	Data Display	5	29	15.1%	3.021
	Total:	**35**	**192**		

As mentioned previously, it is standard practice in the teacher effects literature to base estimates of an individual teacher's ability to increase achievement on students' knowledge of topics that they might have been taught by other teachers. Having data on the correspondence between the classroom curriculum and the assessment curriculum suggests that this concession is a necessity: few teachers—in any nation, in any round of TIMSS—are responsible for teaching 100 percent of the test topics their students were taught. For this reason, among students who have been taught similar percentages of topics, I do not distinguish among them those who were taught exclusively by their test-year teacher. Although this is common practice, I do not consider results from these analyses to be indicative of an individual teacher's effectiveness and do not attempt to calculate a teacher effectiveness score for individual teachers. Instead, results are indicative of the overall instructional effectiveness of each nation's teachers.

Any model of student achievement that attempts to partition variance components to different levels of analysis needs to take into account factors that influence both the assignment of students to classrooms and student achievement (Reardon and Raudenbush 2009). This happens most obviously when schools track students according to ability. In these schools in particular, it seem highly probable that schools determine the assignment of both students and teachers to classrooms based on how well they believe students and teachers will do if assigned to a particular classroom. In effort to eliminate the problem of confounding, I capitalize on information provided in the Principal Questionnaire, which tells us which schools organize math instruction by grouping students by ability. Schools in which principals indicated that student ability is taken into account to organize math *in any manner* are coded 1. All other schools not missing data are coded zero. As best as possible given the data, I also control for student SES, classroom mean SES, classroom mean ability and the variance within classrooms in student ability.

The best indicator of student SES available in the TIMSS fourth-grade population comes from a question posed to students regarding the approximate number of books in their homes. Specifically, students are asked "About how many books are there in your home? (Do not count magazines, newspapers, or your school books.)" and given five response choices, "None or very few (0-10 books)," "Enough to fill one shelf (11-25 books)," "Enough to fill one bookcase (26-100 books)," "Enough to fill two bookcases (101-200 books)," "Enough to fill three or more bookcases (more than 200 books)," which are each accompanied by a graphic illustration of a shelf or bookcase filled with the greatest number of books in that category. (For example, the illustration accompanying the response “Enough to fill one bookcase (26-100 books)” shows a bookcase with 100 books.) TIMSS does not ask fourth-grade students about their parents’ educational background, occupational position or income level because they have found that fourth graders’ knowledge of these facts is weak.[23] I control for student SES by creating two dummy variables indicating high or low SES as whether or not student responses fall one standard deviation above or below the national mean, respectively.

At the student level, I also control for gender (female = 1), immigrant status (student and/or parent(s) are immigrants = 1), age (one standard deviation or more below the national mean = 1) and non-native language (students rarely or never speak the language of the assessment at home = 1) as these factors might influence student achievement or the variance in student achievement within classrooms. For the same reason, at the classroom level I control for the percentage of non-native language speakers in class and class size. In addition to their potential influence on student achievement, I include these variables in the analyses as they might interfere with teachers’ abilities to effectively convey math knowledge to students. At the school level, I control for school size. I include the school size variable for two reasons. First, the TIMSS sample was stratified by school size, with larger schools receiving a higher probability of selection than smaller schools. Second, I include school size because Hong Kong, Japan and Taiwan were more likely to sample two or more classrooms from large schools than they were from small schools.

Unless otherwise specified, all continuous variables are standardized within each nation to have a mean of zero and a standard deviation of one. In other words, for each nation, a score of zero indicates the average score in that nation; means differ across nations. Variables are standardized to ease interpretations. They are centered on the national mean as opposed to the

[23] TIMSS research coordinator, personal communication.

international mean to account for cross-national differences in self-ratings of ability (Lee, Graham and Stevenson 1996) and national differences in response styles (Harzing 2006).

Hong Kong, Japan, Singapore, Taiwan and the U.S. participated in both rounds of the 2003 and 2007 TIMSS fourth-grade population. The U.S. sample includes data from 17,725 students in 994 classrooms in 505 schools. From Hong Kong there are data from 8,399 students in 292 classrooms in 258 schools. The Japan sample contains information from 9,022 students in 339 classrooms in 298 schools, while the Singaporean sample includes data from 11,709 students in 536 classrooms in 359 schools. Rounding out the list is Taiwan, with 8,792 students in 324 classrooms in 300 schools. I run the multilevel analyses on the complete cases (e.g. students not missing information on any variables used in the analyses). This decreases the sample size by 8.9 percent for Hong Kong , 9.0 percent for Japan, 9.3 percent for Singapore, 9.6 percent for Taiwan and 7.4 percent for the U.S.

ARE MATH TEACHERS IN EAST ASIA MORE EFFECTIVE THAN THOSE IN THE U.S.?

Westbury (1992) contends that once curricular differences are taken into account, U.S. math teachers are equally as effective as their Japanese counterparts. At each grade level, Westbury compares just one group of U.S. teachers—those who teach advanced math courses—to Japanese teachers. I conduct this analysis on two subsets of the sample in order to include a broad spectrum of students and teachers from all nations in my analyses. The first subset contains students who were taught 58-63 percent of the test topics, which represents the lower bound of the average percentage of test topics taught in the five nations under investigation. This is the average for Japanese students (and by implication, for Japanese teachers) but is low for the average student (and teacher) in Singapore and the U.S. The number of students in this sample ranges from 116 in Singapore to 1,339 in Japan. In Hong Kong, Taiwan and the U.S. the sample size is 668, 469 and 510, respectively. The second subset includes students who were taught 86-91 percent of the test topics, which represents the upper bound of the averages among nations. This is the average for Singaporean students (and teachers) and just slightly above the U.S. average of 85 percent. This subset contains 837 students from Hong Kong, 224 from Japan, 2,942 from Singapore, 1,223 from Taiwan and 2,042 from the U.S. I use jackknife replicate weights, which take into account the TIMSS sampling design, to calculate accurate point estimates.

Table 3.2 presents the results from these analyses. The table is partitioned into two sections. The top section displays results for the sample of students who were taught 58-63 percent of the test topics; the bottom section displays results for students who were taught 86-91 percent of the test topics. Within each section, nations are ordered from top to bottom by the mean achievement of students in this subsample, which is given in column 2. The top row of each section displays results from the nation with the largest mean achievement score. Columns 3 and 4 indicate the 95 percent confidence interval for mean achievement. The last column provides insight into relative instructional effectiveness across nations; it tells us the average points earned for every percentage of test topics taught.

The results displayed in Table 3.2 do not support Westbury's (1992) earlier findings. In both subsamples, mean achievement in the U.S. is outside the 95 percent confidence interval of mean achievement in every East Asian nation. The mean achievement of students who were taught 58-63 percent of the test topics in each nation ranges from a high of 586 in Hong Kong to a low of 518 in the U.S.—a difference of 68 points or about two-thirds of a standard deviation. Among students who were taught 86-91 percent of the test topics, mean achievement within each nation ranges from a high of 597 in Singapore, to a low of 530 in the U.S. Again, the difference between the two nations is a substantial 67 points.

Table 3.2 Cross-National Differences in Average Math Achievement

	Math Achievement			Points per Topic
		95% confidence interval		
Nation	Mean			Mean
Among students who were taught 58-63% of the test topics				
Hong Kong	586.2	579.7	592.7	9.65
Singapore	573.7	564.7	582.7	9.26
Japan	569.1	565.0	573.2	9.40
Taiwan	562.6	557.7	567.6	9.30
United States	518.2	513.4	523.0	8.58
Among students who were taught 86-91% of the test topics				
Singapore	597.0	592.1	601.8	6.72
Hong Kong	591.3	586.2	596.3	6.72
Taiwan	570.8	567.1	574.6	6.42
Japan	561.1	554.8	567.4	6.34
United States	529.9	525.9	533.9	6.01

The figures displayed in column 4 of Table 3.2 describe cross-national differences in the rate at which teachers increase achievement. It is based on a simple measure of productivity and is calculated by dividing student achievement (the output) by the assessment-weighted percentage of test topics students were taught (the input). The results tell us the average number of achievement points teachers in each nation produce for every one percent of test topics they teach, or the points per topic taught. In the top section of Table 3.2, we see that teachers in Hong Kong tend to increase student achievement by 9.65 points per topic. By contrast, U.S. teachers increase achievement at the rate of 8.58 points. The difference in points per topic between Hong Kong and the U.S. (9.65-8.58=1.07) is three times larger than the difference between Hong Kong and Singapore (9.65-9.3=0.35), the least-highest achieving East Asian nation in that subsample. In the 86-91 percent subsample, Singaporean and Hong Kong teachers appear to be the most effective; they increase achievement by an average of 6.7 points per topic. In this subsample, too, the figure for U.S. teachers is much lower (6.01 points per topic) than their East Asian counterparts.

To gain a sense of the relationship between the quantity of instruction and student achievement, I compare each nation's results across the subsamples. According to previous international assessment results (Medrich and Griffith 1992) as well as Sørensen and Morgan (2000), student achievement should increase with quantity of instruction. If this were true, then we should see that within each nation, achievement is higher among students who learned 86-91 percent of the test topics compared to students who learned 58-63 percent. The results displayed in Table 3.2 provide mixed support for this argument. In Singapore, Taiwan and the U.S., average student achievement is higher in the 86-91 percent subsample than in the 58-63 percent subsample. This pattern is not evident in the sample data from Hong Kong and Japan, however. In Hong Kong, mean achievement is higher among the 86-91 percent subsample, but it is within the 95 percent confidence interval of achievement in the 58-63 percent subsample. In Japan, the pattern contradicts previous findings: average achievement is actually lower in the 86-91 percent subsample than in the 58-61 percent one, suggesting the possibility that achievement in Japan *declines* as the quantity of instruction increases.

It is possible that the positive relationship between quantity of instruction and student achievement that is evident in Singapore, Taiwan and the U.S. exists because in those nations, quality of instruction is correlated with quantity of instruction. In other words, teachers who can teach really well tend to cover more material. The analyses conducted so far do not permit us to examine the relative impact of quality of instruction versus quantity of instruction on student achievement. To address this research question, in the next section I build three-level hierarchical linear models that allow me to assess, 1) the extent to which teachers within nations vary in their effectiveness, 2) whether or not teachers who teach more material produce more learning and, 3) the degree to which the association between quantity of instruction and student achievement differs across nations.

DOES TEACHER EFFECTIVENESS VARY LESS IN EAST ASIA?

International comparative educational research emphasizes the role that structural features of East Asian educational systems play in promoting high levels of teacher effectiveness. These structural features, such as robust teacher education and development programs (Stigler and Hiebert 1999; Wang et al. 2003), should not only increase the overall levels of teacher effectiveness, but also should decrease the amount of variation in teacher effectiveness. Moreover, if quantity of instruction determines student achievement, as Westbury (1992) and Sørensen and Morgan (2000) contend, then I should find higher levels of achievement in classrooms that cover more test topics. Since curriculum coverage can also vary across classrooms within the same school (especially in those that track by ability), including a measure of quantity of instruction might also explain some of the variation in teacher effectiveness within each nation.

I develop a three-level hierarchical model in which students are nested within classrooms within schools to empirically assess these hypotheses. To take into account the possibility that the TIMSS math assessment is not a good indicator of the material that teachers covered in class, and would thus provide biased estimates of teacher effects (Martineau 2006), I carry out the analyses twice, once on all complete cases and once again on the subset for whom the TIMSS test curriculum seems to correspond well to their classroom curriculum. I use the same models for both sets of analyses.

I first examine a model that predicts student achievement using only an indicator of test round (whether or not data come from the 2003 round). In the second model, I include classroom and school contextual features that might influence student achievement or the assignment of students and teachers to classrooms within schools. In the third model, I include the quantity of instruction variable.

Statistical model

The complete level-1 or student-level model is:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(2003)_{ijk} + \pi_{2jk}(Ability)_{ijk} + \pi_{3jk}(Low\ SES)_{ijk} + \pi_{4jk}(High\ SES)_{ijk} + \pi_{5jk}(Age)_{ijk} + \pi_{6jk}(Language)_{ijk} + \pi_{7jk}(Immigrant)_{ijk} + e_{ijk}$$

where Y_{ijk} is student *i* in the *j* classroom of the *k*th school's achievement score and e_{ijk} is a random "student" effect that captures the deviation of student *ijk*'s score from the classroom mean (Raudenbush and Bryk 2002). I assume that these effects are normally distributed with a mean of 0 and variance σ^2.

Mean student achievement (π_{0jk}) is the only coefficient allowed to vary over classrooms and schools. All other coefficients are fixed because how they might vary over classrooms and schools is not the central focus of this chapter. Mean student achievement is the outcome variable of the level-2 or classroom model, which adjusts predictions of mean student achievement at the classroom level by classroom contextual factors in the second analysis model. The third model also adjusts for the quantity of instruction students received. The classroom contextual variables used are the percentage of non-native language speakers in class, classroom mean prior ability, variance in prior ability in class, classroom mean SES and class size. The complete level-2 or classroom model is:

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(\%\ Lang.)_{jk} + \beta_{02k}(Ability)_{jk} + \beta_{03k}(SD\ Ability)_{jk} + \beta_{04k}(SES)_{jk} + \beta_{05k}(Size)_{jk} + \beta_{06k}(Quantity)_{jk} + r_{0jk}$$

where r_{0jk} is a classroom-specific random effect that tells us how much the *jk* classroom's mean achievement deviates from the school mean. The variance of r_{0jk} or τ_π tells us how much achievement varies across classrooms net of the student and classroom level controls. I assume that it is normally distributed with a mean of 0.

The level-3 model is the school model. As in the classroom model, only achievement is allowed to vary across schools. I adjust estimates of school mean student achievement by a dummy variable indicating that schools group by ability and an indicator of school size. The complete level-3 model is:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(Ability\ Group)_k + \gamma_{002}(Size)_k + u_{ook}$$

where u_{00k} is a school-specific random effect with mean 0 and variance τ_β, and γ_{001} and γ_{002} give us the effects of ability grouping and school size, respectively, on school mean achievement, $\beta_{00k.}$. All level-1 coefficients are constrained to be constant within classrooms and schools, which means that $\pi_{1jk} = \beta_{10k} = \Upsilon_{100}$, $\pi_{2jk} = \beta_{20k} = \Upsilon_{200}$ ... $\pi_{7jk} = \beta_{70k} = \Upsilon_{700}$. Similarly, all classroom-level variables are fixed within schools such that $\beta_{01k} = \Upsilon_{010}$, $\beta_{02k} = \Upsilon_{020}$ ... $\beta_{05k} = \Upsilon_{050}$.

Table 3.3 displays the estimates and standard errors for average classroom achievement (π_{0jk}) and quantity of instruction (β_{05k}) as well as the random effects estimates σ^2, τ_π, and τ_β and the percentage of the total variation in achievement status that is explained by each level. The number of students, schools and classrooms in the sample, as well as the average number of students per classroom, are also included in Table 3.3. Results from analyses on the Complete Case sample are on the right, results from the Test-Appropriate sample are on the left. For space reasons, results from the Null model are not included in the Table. Columns 2-3 and 6-7 display results from the models that include classroom and school contextual variables and columns 4-5 and 8-9 display results from the model that includes quantity of instruction. Each horizontal panel of Table 3.3 displays the results for a different nation.

How much do math teachers within each nation vary in effectiveness?

Starting with the first panel of Table 3.3, which displays results from the hierarchical linear model analysis of math achievement status in the U.S., we see that estimates of the between-classroom variance component account for around 11.5 percent of the variation in student achievement status in both subsamples of data. This is consistent with previous studies (Nye et al. 2004). According to the second panel of Table 3.3, which display results from the analyses using data from Hong Kong, estimates of the between-classroom variance component explain between 16-19 percent of the variation in student achievement status. In Japan (the third panel of data in Table 3.3), estimates of the between-classroom variance component account for around 2 percent of the variation in student achievement. Estimates of the between- classroom variance component from the Singaporean data (presented in the fourth panel of Table 3.3) explain 20-22 percent of the variation in student achievement. In Taiwan, estimates of the between-classroom variance component account for 2-3 percent of the variation in student achievement status.

Table 3.3 Results from Three-level Models of Student Achievement

Nations	Complete Case sample				Test-Appropriate sample			
	Model 1		Model 2		Model 1		Model 2	
United States								
Fixed Effects	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*
Intercept	536.8	2.19	533.8	2.25	537.3	2.17	532.8	2.3
Quantity of Instruction			0.32	0.28			0.37	0.33
Random Effects	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*
Students e_{ijk}	3188.4	81.1%	3190.1	81.2%	3193.3	81.2%	3193.3	81.6%
Classes r_{0jk}	446.5	11.4%	445.8	11.4%	454.4	11.6%	445.7	11.4%
Schools u_{00k}	298.5	7.6%	291.1	7.4%	285.1	7.2%	275.8	7.0%
N students	11,815		11,815		10,967		10,967	
N schools	424		424		406		406	
N classes	730		730		681		681	
Class Avg.	16		16		16		16	
Hong Kong								
Fixed Effects	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*
Intercept	599.1	3.11	598.5	3.21	596.6	3.61	596.5	3.53
Quantity of Instruction			0.03	0.38			0.04	0.41
Random Effects	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*
Students e_{ijk}	2487.9	81.3%	2496.3	80.7%	2503.6	81.7%	2503.5	81.7%
Classes r_{0jk}	564.9	18.5%	564.5	18.3%	504.1	16.5%	506.5	16.5%
Schools u_{00k}	8.2	0.3%	32.0	1.0%	56.5	1.8%	54.2	1.8%
N students	6,738		6,738		6,120		6,120	
N schools	223		223		206		206	
N classes	248		248		227		227	
Class Avg.	27		27		27		27	
Japan								
Fixed Effects	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*
Intercept	569.4	3.10	570.8	3.35	569.2	4.36	565.4	2.89
Quantity of Instruction			0.10	0.32			-0.44	0.45
Random Effects	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*
Students e_{ijk}	4143.7	94.6%	4146.8	94.3%	4047.3	94.9%	4059.8	95.2%
Classes r_{0jk}	154.6	1.9%	89.3	2.0%	109.3	2.6%	98.6	2.3%
Schools u_{00k}	83.1	3.5%	162.4	3.7%	110.2	2.6%	105.5	2.5%
N students	7,141		7,141		3,656		3,656	
N schools	265		265		135		135	
N classes	299		299		154		154	
Class Avg.	24		24		24		24	

Table 3.3 Continued

	Complete Case sample				Test-Appropriate sample			
Nations	**Model 1**		**Model 2**		**Model 1**		**Model 2**	
Singapore								
Fixed Effects	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*
Intercept	583.9	8.62	584.7	8.63	583.7	8.64	584.1	8.63
Quantity of Instruction			-0.08	0.42			-0.03	0.42
Random Effects	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*
Students e_{ijk}	2775.4	76.8%	2777.0	76.3%	2776.5	76.8%	2776.4	76.8%
Classes r_{0jk}	775.1	21.5%	782.5	21.5%	745.0	20.6%	744.72	20.6%
Schools u_{00k}	62.8	1.7%	79.4	2.2%	91.8	2.5%	92.69	2.6%
N students	9,627		9,627		9,613		9,613	
N schools	323		323		323		323	
N classes	459		459		458		458	
Class Avg.	21		21		21		21	
Taiwan								
Fixed Effects	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*	*coef.*	*se*
Intercept	593.3	2.23	593.1	2.22	593.0	2.24	592.1	2.19
Quantity of Instruction			0.17	0.32			0.24	0.31
Random Effects	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*	*Var.*	*%*
Students e_{ijk}	2838.4	95.1%	2840.5	94.9%	2828.6	95.1%	2828.5	95.3%
Classes r_{0jk}	65.8	2.2%	77.4	2.6%	50.72	1.7%	83.44	1.9%
Schools u_{00k}	81.2	2.7%	74.9	2.5%	96.33	3.2%	57.23	2.8%
N students	7,458		7,458		7,073			7,073
N schools	264		264		250			250
N classes	284		284		268			268
Class Avg.	26		26		26			26

To gauge the magnitude of the difference in achievement status between having a highly effective teacher versus a less effective teacher in each nation, I calculate the difference between the minimum and maximum mean classroom achievement scores that fall in the 95 percent plausible values range of the standard normal distribution. For each nation, I measure this difference using estimates from the Quantity of Instruction Models. Calculations based on the Complete Case sample are given first, followed by estimates from the Test-Appropriate sample.

Results from these computations suggest that the difference between being assigned to a highly effective versus a less effective teacher in the U.S. ranges from 79 to 83 achievement points, or four-fifths of a standard deviation in achievement scores. In Hong Kong, the difference between being assigned to a highly effective teacher versus a less effective teacher is about the same as it is in the U.S.: the difference ranges from 82 to 88 points. In Japan, the difference between being assigned to a highly effective teacher versus a less effective teacher is either 37 or 39 points, depending on the sample data used. In Singapore, the difference between being assigned to a highly effective teacher versus a less effective teacher is about 110 points in either estimation. In Taiwan, the difference between being assigned to a highly effective teacher versus a less effective teacher is 35 or 30 points, depending on the sample.

We can also use the range of plausible values among teachers within each nation to gain insight into how the most effective teachers in the U.S. stack up to their counterparts in East Asia. For simplicity, I will limit the discussion to results from the model including the quantity of instruction from the Test Appropriate sample (the results displayed in the final two columns of Table 3.3). The mean classroom achievement of highly effective U.S. teachers is 574 ($=532.8+(1.96*\sqrt{445.7})$). This figure is just slightly higher than the average classroom mean achievement score in Japan (565.4, falling within a 95 percent confidence interval of 560 to 571), suggesting that, when the sample is limited to students who learned between 60 and 100 percent of the TIMSS test topics, the most effective teachers in the U.S. rank roughly on par with a teacher of average effectiveness in Japan. Compared to the rest of the East Asian nations, however, the most effective math teachers in the U.S. are not as effective as the average. But, the most effective teachers in the U.S. are more effective than the least effective teachers in Hong Kong and Singapore, where the average mean achievement score in the bottom 5 percent of classrooms in those nations comes to 552 (=596.5-(1.96*√506.5)) and 531 (=584.1-(1.96*√744.7)), respectively. In Taiwan, the least effective teachers who covered between 60 and 100 percent of the TIMSS topics appear to be equally as effective as the most effective teachers in the U.S. (The average mean achievement score in the bottom 5 percent of classrooms in Taiwan is 577, compared to the average mean achievement score of 574 in the 95 percent of classrooms in the U.S.)

These findings indicate that teacher effectiveness does not vary more within the U.S. than it does in all East Asian nations. Teacher effectiveness does vary less in Japan and Taiwan than it does in the U.S. but there is considerably more variation in teacher effectiveness in Singapore. While differences between classrooms within schools explains more of the variation in student achievement in Hong Kong than in the U.S., estimated differences between having a highly effective teacher versus a less effective teacher are roughly equivalent in the two nations.

Do teachers who teach more material produce more learning?

The coefficients and standard errors for the quantity of instruction variables are given in the third row of every panel in Table 3.3. The results from all samples and nations indicate that in these data, the relationship between quantity of instruction and achievement is weak. The coefficient is never greater than 0.37 (in the Test-Appropriate Sample for the U.S.) and is negative in Japan and Singapore. Further, the relationship between quantity of instruction and student achievement is not statistically significant in any nation or sample. The relation between quantity of instruction and student achievement is similar across nations in that it is weak and statistically insignificant in all models and samples for each nation.

Results from the models using the Test-Appropriate sample data (the right-hand side of Table 3.3) indicate that the addition of the quantity of instruction variable in the final model did not affect estimates of the between-classroom variance component in any nation, which suggests that differences across classrooms within schools in the quantity of instruction teachers provide do not explain variations in teacher effectiveness.

Comparing between-classroom variance component estimates from the Complete Case sample to those from the Test-Appropriate sample sheds light on the effect that relying on a "general use" assessment has on estimates of teacher effectiveness. In each nation, estimates of the between-classroom variance component vary slightly across subsamples. The greatest difference appears in Hong Kong, where the percentage of variance explained by between-classroom differences declines from 18.5 percent in the Complete Case sample to 16.5 percent in the Test-Appropriate sample. In the U.S., the difference across subsamples in between-classroom variance component estimates is quite small—0.2 percent, suggesting that the correspondence (or lack thereof) between the test curriculum and the classroom curriculum does not strongly influence estimates of teacher effects.

CONCLUSION

For several decades researchers have claimed that East Asian math teachers are more effective than their U.S. counterparts. However, there is limited empirical evidence to support or repudiate this argument. The analyses presented in this chapter contribute much-needed empirical insight into international differences in variations in teacher effectiveness across nations. Using pooled data from the 2003 and 2007 rounds of TIMSS from Hong Kong, Japan, Singapore, Taiwan and the U.S., I show how teacher effectiveness in the fourth grade varies within and across nations. I also test hypotheses drawn from the international comparative literature as well as from U.S.-based research to assess the relationship between quantity of instruction and student achievement.

Results from this study suggest that math teachers in East Asian nations are indeed more effective than their U.S. counterparts. I show that, accounting for student and classroom characteristics, among students who have received the same quantity of instruction, average achievement is higher in East Asian nations than in the U.S. and further, that the differences in average achievement are not insignificant. These findings contradict Westbury's (1992) research, which shows that U.S. and Japanese eighth graders who were taught similar percentages of algebra topics and scored similarly on a pre-test scored equally as well on the algebra section of the SIMS post-test.

There are several possible explanations for the inconsistency between the findings presented in this chapter and Westbury's (1992). These involve differences in methodology and time period. Westbury's sample of U.S. teachers contains only those who were assigned to teach advanced math courses in the U.S. Given that advanced math courses in the 1980s in particular were usually only offered at high-performing schools (Gamoran 1986), it is likely that this select sample is composed of the most effective eighth-grade math teachers in the U.S. The subsamples of U.S. teachers that I use are far less likely to include only highly effective teachers. Additionally, extensive curricular reforms have occurred in both the U.S. and Japan in the three decades since the data Westbury used were collected. For example, U.S. math curriculum is far more extensive at all grade levels than it was in the early 1980s. It is possible that this ever-changing curricular environment has influenced U.S. math teachers' ability to promote achievement, which would also explain the inconsistencies between findings.

The findings of this chapter suggest that the relationship between quantity of instruction and student achievement is much more complex than the idea espoused by earlier international comparative scholars as well as Sørensen and Morgan (2000), which is: "the more content students are taught, the more they learn, and the better they perform on the achievement tests" (Medrich and Griffith 1992: 30). I find a very weak and statistically insignificant relationship between quantity of instruction and student achievement in every nation under investigation. It is likely that the relationship between quantity of instruction and student achievement is not clear because a number of previously overlooked factors, such as curriculum sequencing and instructional strategies, are interfering in the relationship. It is possible that previous international assessments provide evidence of a positive relationship between quantity of instruction and student achievement because at that time, the educational systems that covered more content also had more effective curricular guidelines and/or more effective teachers. In the intervening years, as more nations—including the U.S.—strove to provide greater content coverage, the association between quantity of instruction and curriculum sequencing, guidelines, instructional strategies, etc., weakened thereby allowing us to see that quantity of instruction alone does not determine student achievement. Future research might attempt to assess the relative importance of factors such as curriculum sequencing and curricular guidelines on the effectiveness of instruction.

Results from this chapter also dispute the hypothesis that the difference between effective and less effective teachers in the U.S. is greater than it is in East Asian nations. Differences between classrooms in the U.S. account for about 11 percent of the variation in achievement, which mirrors Nye et al.'s (2004) findings for achievement status in math. While teachers in Japan and Taiwan vary less in their effectiveness, teachers in Hong Kong and Singapore vary more in their ability to promote achievement than math teachers in the U.S.

The sizeable differences within East Asia in terms of how much teachers vary in their ability to promote achievement calls into question the suggestion that structural features of East Asian educational systems act to reduce the amount of variation in teacher effectiveness in those nations. The findings from this chapter's analysis suggest that this argument holds for the cases of Japan and Taiwan, but not for Hong Kong or Singapore. In Japan and Taiwan, teachers tend not to differ much in terms of effectiveness. Differences between classrooms and schools combined account for just 6 percent of the variation in student achievement in the final models. Coleman et al. (1966) used a similar finding to argue that families tend to matter more than schools for student achievement. However, Sørensen and Morgan (2000) provide a difference take on the issue. They point out that if schools and teachers provide all students with the same learning opportunities, then students with more natural ability, and those who put more effort into their learning (or both), will outperform their peers. In other words, student characteristics *should* explain the largest proportion of variance in student achievement if the educational system ensures that all students receive the same quantity and quality of instruction. Thus, the findings from Japan and Taiwan could imply that it matters very little which school or teacher a student is assigned to because *all* schools and teachers in Japan and Taiwan provide an *equal* amount of high-quality learning opportunities. It is what students do with those opportunities that determines their achievement.

Compared to Japan, Taiwan and the U.S., the amount of variation in teacher effectiveness within Hong Kong and Singapore is substantial. In both nations, however, differences between schools account for very little of the variation in student achievement in the final models. In other words, in Hong Kong and Singapore, which teacher a student is assigned to matters more than the school a student attends. It might be that all schools in Hong Kong and Singapore provide the same amount of resources to their teachers such that differences in teachers' skills are intensified. Why this situation occurs in Hong Kong and Singapore but not in Taiwan and Japan is an interesting question. Future research might examine similarities and differences between Japan and Taiwan compared to Hong Kong and Singapore in the amount of autonomy given to teachers and in the frequency and depth of professional support they receive.

Chapter 4

Parents

THIS CHAPTER SHIFTS FOCUS from the relationship between math teachers and student achievement to the relationship between schools and parents. Parents play a principal role in their children's early educational development. For this reason, it is possible that a disparity between East Asia and the U.S. in the proportion of parents who are effectively involved in their children's education explains some of the East Asian-U.S. achievement gap. Indeed, it is not uncommon for laypeople and scholars, including Stevenson and Stigler (1992), to credit parents in East Asia for their children's superior performance on international assessments. Despite the widespread belief (fueled most recently by Chua's 2011 best-selling *Battle Hymn of the Tiger Mother*) that East Asian parents are relentless in their quest to raise academic overachievers, there is no study of nationally representative data to support this argument. Further, while it is commonly believed among policymakers that it is possible to increase the proportion of parents who are effectively involved in their children's education (doing so is now a major policy goal at the federal, state and local levels in the U.S. (Domina 2005) as well as in Hong Kong (Cheung 2001), Singapore (Singapore Ministry of Education 2012) and Taiwan (Hung 2007)), the prevailing theory on the East Asian-U.S. parental involvement gap would suggest that such policy efforts, at least in the U.S., are futile.

A sociological approach that is both empirically and theoretically grounded could contribute to this debate. Although some sociologists, notably Coleman (1991) and Epstein (1986, 1995), have put forth theories to explain how the proportion of parents who are involved in their children's education might be increased, their theories have yet to be applied to East Asian-U.S. comparative research. In this chapter, I discuss the limitations of the existing evidence and prevailing theory on the East Asian-U.S. parental involvement gap before providing an alternative theory using Coleman and Epstein's work. Unfortunately, it is not possible with the TIMSS data (nor with any other internationally comparable dataset on early education) to isolate the mechanisms through which schools encourage parents to become involved. This is because no internationally comparable dataset on elementary-school education focuses specifically on the strategies schools utilize to encourage parents to become involved in their children's education, nor includes multiple, suitable indicators of parental involvement from the U.S. and the East Asian nations under investigation.[24]

I conclude this chapter by exploring the relationships between features of the school community that Coleman and Epstein link to parental involvement and the likelihood of parental involvement within each nation, and by assessing how these relationships vary across nations. Unfortunately, given the state of the evidence, it would be premature to draw any firm conclusions; my aim is to put the study of the East Asian-U.S. parental involvement gap on a sounder footing so that such conclusions might be made possible.

[24] The Progress in Reading Literacy Study (PIRLS) is the only internationally comparable dataset to survey the parents of elementary-school children and thus to provide more insight into parental involvement within each nation. However, Japan and Taiwan do not participate in PIRLS and responses to the family survey are not available from the U.S.

PARENTAL INVOLVEMENT IN CHILDREN'S EDUCATION

A necessary precondition for this research is a thorough definition of the concept of parental involvement (Feuerstein 2000). The concept has been defined broadly to encompass all parenting practices that might influence children's educational development. Epstein (1995) created a typology to account for the concept's multifaceted nature. She identifies six different types of parental involvement: 1) child-rearing practices that establish the home as an environment conducive to study, including practices indirectly related to educational development, such as nutrition and health, 2) home-school communications concerning school programs and children's progress, 3) volunteering at school, 4) home-based activities related to school including helping with homework, course selection and planning, 5) participation in school decisions through active parent-teacher associations (PTA), advisory councils or special committees, and 6) collaborating with the community to strengthen school programs and children's learning opportunities.

There does not appear to be a consensus on the form of parental involvement that best captures parents' commitment to their children's education. Over the years, scholars have relied on a variety of indicators to measure parental involvement, including: parents' expectations for children's educational attainment (Coleman 1988; Roksa and Potter 2011; Sandefur et al 2006), parent-child discussions about school (Desimone 1999; Kao 2004; Teachman, Paasch and Carver 1996), children's participation in high-status cultural activities (Dumais 2006), and the number of social ties that parents have to the parents of their children's friends (Carbonaro 1998; Morgan and Sørensen 1999). The multiplicity of ways in which parents can be involved in their children's education complicates efforts to compare relative levels of parental involvement across nations using limited, survey-based information—a topic I return to in the next section.

There are two additional points to note regarding parental involvement. First, although much of the sociological work is narrowly focused on the connections between parental involvement and children's achievement scores or grades, Epstein (1986, 1995) and others (Hoover-Dempsey et al 2005; Domina 2005) suggest that many forms of parental involvement influence learning indirectly, through their effect on children's non-cognitive attributes such as feelings of self-efficacy, organizational skills and engagement with school. Thus, parental involvement can be considered effective if it is associated with either cognitive or non-cognitive attributes. With this in mind, I test the relevance of my measure of parental involvement by assessing the relations between it and student's math achievement as well as students' engagement with school and self-confidence in learning math. Second, U.S.-based research has found that both the extent and efficacy of parental involvement might decline with children's age (Domina 2005). For this reason, analyses that rely on data from elementary-school students as opposed to high-school students should be more fruitful—which is why I chose to rely solely on data from the TIMSS fourth-grade populations for these analyses.

EXISTING EVIDENCE OF AN EAST ASIAN-U.S. PARENTAL INVOLVEMENT GAP

The best source of evidence of a parental involvement gap between East Asia and the U.S. comes from a series of studies conducted by Stevenson and colleagues during the 1980s and 1990s in Beijing, Chicago, Minneapolis, Sendai and Taipei.[25] Although Stevenson and colleagues (Chen and Stevenson 1989; Stevenson and Stigler 1992; Stevenson, Chen and Lee 1993) clearly state that the level of parental involvement is lower in the U.S. than in East Asian nations, they do not clearly define whether this is because a smaller proportion of parents in the U.S. than in East

[25] Much of the evidence of an East Asian-U.S. parental involvement gap comes from studies of immigrants to the U.S. Given that Asian immigrants to the U.S. are a select group and are thus not representative of their native countrymen, I do not draw on evidence from immigrant studies in this chapter.

Asian nations is involved in their children's education or because, compared to their counterparts in East Asia, U.S. parents are less *effectively* involved in their children's education. In this section, I show why the existing evidence does not necessarily support the second argument (that East Asian parents are more effectively involved), although it does provide some compelling support for the first argument (that a greater proportion of parents in East Asian nations than in the U.S. is involved).

One of the key findings from Stevenson and colleagues' research is that U.S. parents tend to be more satisfied than parents in East Asia with their children's academic achievement and with the quality of their children's schools. As Stevenson, Chen and Lee (1993) explain, this trend continued throughout the 1980s—despite a growing awareness among the U.S. parents they interviewed that the U.S. educational system lags behind those of other industrialized nations. Stevenson et al's (1993) study shows that the U.S. mothers interviewed in 1990 were just as likely as the U.S. mothers interviewed in 1980 and 1984 to describe their attitude towards their children's academic achievement as "very satisfied" (as opposed to "satisfied" or "not satisfied"). In each survey round, more than 40 percent of U.S. mothers surveyed gave this response, compared to less than 10 percent of mothers in Sendai and Taipei. Similarly, when asked to rate the efficacy of their children's schools, almost twice as many U.S. mothers than Japanese or Taiwanese mothers responded that their children's schools were doing an "excellent" or "good" job. According to Stevenson et al, the fact that U.S. mothers are cognizant that the U.S. lags behind other nations on international assessments but overall remain more satisfied with their children's education than Japanese and Taiwanese mothers signals their low academic standards. Similarly, the authors take as a sign of higher academic standards the fact that mothers in Japan and Taiwan continue to express a low level of satisfaction with their children's education despite their awareness of their nation's stellar performance on international assessments. The authors use this cross-regional disparity in levels of satisfaction to argue that U.S. parents are less effectively involved than their counterparts in East Asia.

However, the fact Stevenson, Chen and Lee (1993) found that the groups of U.S. mothers surveyed tend to be more satisfied with their children's education than the groups of mothers surveyed in Japan and Taiwan could have more to do with national differences in response styles than with real differences in mothers' attitudes. Studies focused on international differences in response styles consistently show that, when asked to rank their responses on a scale, East Asians tend to check the middle response categories while respondents in the U.S. tend to check the extreme categories (Harzing 2006). This could explain why Chen and Stevenson (1989) find average ranked responses to be higher among U.S. mothers than among Japanese and Taiwanese mothers. Given that there are national differences in response styles, the evidence that U.S. parents are less effectively involved in their children's education than East Asian parents would be more convincing if researchers were to either factor national differences in response styles into their analyses or to rely on evidence that is not based on parents' rated responses to survey questions.

Additional evidence that U.S. parents are less effectively involved than parents in East Asia is that, of the mothers surveyed in 1980 and 1986, those in Chicago and Minneapolis spent less time, on average, helping their children with homework each week than parents in Beijing, Sendai and Taipei (Chen and Stevenson 1989). Yet the amount of time that parents spend helping their children with homework might not act as a good indicator of parents' commitment of resources to their children's education because it strongly depends on the amount of homework that their children's teachers assign. For example, if teachers do not assign any math homework, then even parents who undertake other forms of involvement in their children's education cannot spend time helping with math homework. Given that math teachers in East Asian nations tend to assign more homework than their counterparts in the U.S., it is possible that a disparity in minutes of homework assigned (as opposed to a difference in the effectiveness of parents' involvement) explains Chen and Stevenson's findings. To test this hypothesis, I use data from Chen and Stevenson (pp. 556, 558) to calculate the ratio of average minutes mothers spent helping with homework to average minutes of homework teachers assigned per week. The resulting ratios indicate that—almost without exception—mothers in Chicago and Minneapolis actually spent *more time*, relative to the amount of homework assigned, helping with homework than did mothers in any Asian city.

Table 4.1 Ratio of Minutes of Homework Assigned to Minutes of Help Mothers Provide

	1980 Study		1986 Study		
Cities	Grade 1	Grade 5	Grade 1	Grade 3	Grade 5
Chicago	.	.	2.87	0.99	0.78
Minneapolis	4.57	0.71	.	.	.
Beijing	.	.	2.25	0.90	0.58
Taipei	0.68	0.35	2.31	.	0.92
Sendai	1.52	0.39	1.85	.	0.63

Table 4.1 displays the ratios of average minutes of homework help to average minutes of homework assigned that I calculated using data from Chen and Stevenson (1989). The table is divided into 2 panels. The panel on the left displays the ratios derived from their 1980 data, which include Grades 1 and 5 in Minneapolis, Taipei and Sendai. The panel on the right displays ratios derived from Chen and Stevenson's 1986 data, which includes Grades 1, 3 and 5 in Chicago and Beijing and Grades 1 and 5 in Taipei and Sendai.

The ratios presented in Table 4.1 for mothers of Grade 1 students indicate that for every minute of homework assigned each week, mothers in Minneapolis and Chicago spent an average of four-and-a-half minutes and almost three minutes, respectively, helping with homework. In comparison, for every minute of homework assigned, mothers of first graders in Beijing spent an average of two-and-a-quarter minutes helping with homework while across survey rounds mothers in Sendai spent less than two minutes, on average, helping with homework for every minute of homework assigned and mothers in Taipei spent about forty-five seconds in 1980 and two-and-a-half minutes in 1986 helping with homework for every minute of homework assigned. The only instance in which mothers in the U.S., on average, did not spend more time helping with homework relative to the minutes of homework assigned is in the 1986 study of mothers of Grade 5 students. Here, the ratio is highest in Taipei (0.92) followed by Chicago (0.78), Sendai (0.63) and Beijing (0.58).

Once the average minutes of homework that teachers assign each week is taken into account, the evidence used to support the idea that parents in East Asia are more involved in their children's education than parents in the U.S. now appears to suggest the opposite: that parents in the U.S. tend to be more involved than their counterparts in East Asia. However, there is another contextual factor that likely influences this measure of parental involvement. This is the amount of time that children need to complete their homework assignments. Mothers in the U.S. might spend more time helping their children with homework relative to the minutes of homework assigned because their children need more help than do children in East Asia. Chen and Stevenson (1989: 559) allude to this possibility when they note that the amount of time mothers spent helping with homework is negatively correlated with 24 of their 27 measures of cognitive ability. In other words, in all three nations, mothers whose children are struggling academically tend to invest more time helping their children with homework. Given that the amount of time parents spend helping with homework appears to be strongly influenced both by the amount of homework assigned and by children's cognitive attributes, cross-national comparisons of the time mothers spend on homework help would yield compelling evidence of a parental involvement gap only if the analysis pool were narrowed to the parents of students who are struggling academically and the amount of homework assigned were taken into account. Even then it would be difficult, without longitudinal data on students' cognitive and non-cognitive attributes, to use time spent on homework help to assess whether parents in the U.S. are less effectively involved than parents in East Asian nations.

So far in this section I have given three examples of the evidence Stevenson and colleagues (Chen and Stevenson 1989; Stevenson and Stigler 1992; Stevenson, Chen and Lee 1993) use to buttress their argument that parents in East Asian nations are more effectively involved in their children's education than parents in the U.S. I have suggested that the evidence based on ranked responses is biased by national differences in response styles. I have also shown how overlooked contextual factors weaken the evidence. But the greatest limitation to the existing evidence is that it does not demonstrate cross-regional differences in the effectiveness of parents' involvement. To do so, research would need to show that the effect of parental involvement on growth in student achievement is greater in East Asian nations than in the U.S. holding other factors that influence achievement constant. Scholarship has yet to provide this type of analysis. For these reasons, there is little evidence to suggest that U.S. parents are less effectively involved in their children's education than parents in East Asian nations.

The best evidence of an East Asian-U.S. parental involvement gap comes from Stevenson and Stigler (1992), who show that even though homes are considerably smaller, on average, in

Japan and Taiwan than in the U.S., 98 percent of the parents interviewed in Sendai and 95 percent of parents in Taipei had provided their fifth graders with a study desk at home compared to just 63 percent of parents in Minneapolis. It is unlikely that this finding reflects a greater household income disparity in the U.S. than in Japan and Taiwan: according to Deininger and Squire (1996), the average Gini coefficient of income inequality over the latter half of the 20th century is roughly equivalent between the U.S. and Japan (35.3 and 34.8, respectively) and considerably smaller in Taiwan (29.6) than in Japan. The fact that almost all parents Stevenson and colleagues interviewed in Japan and Taiwan had provided their fifth-graders with a study desk at home compared to less than two-thirds of parents interviewed in the U.S. gives reason to believe that a greater proportion of parents in East Asian nations than in the U.S. is involved in their children's education.

There is some merit to measuring parental involvement as provision of a desk at home. Provision of a desk at home is a permanent indicator of parents' commitment of resources to their children's education, as opposed to a form of involvement that parents might engage in only sporadically. Further, there are not too many reasons why parents would provide their child with a desk other than to benefit children's schoolwork.[26] It is also a common way that parents can demonstrate the importance of education to their children in both East Asia and the U.S. Another advantage is that every student should benefit from having a place to study at home regardless of their individual abilities and needs. Even if educators point out to parents the benefits of providing a study space, and point out that a suitable desk can be had at all price points, whether

[26] The fact that there are not too many reasons why parents would provide their child with a desk at home other than their belief that it is important for their child to have a place to study is noteworthy because studies suggest that the reasons why parents engage in certain forms of involvement vary across racial/ethnic groups. For example, using data from the 1988 NELS, Kao (2004) shows that the relationship between parent-child discussions about school and student achievement are positive among Hispanic immigrants and negative among Asian immigrants. She hypothesizes that this is because Asian parents only discuss school on a regular basis when their children are performing poorly.

or not parents decide to provide the desk is a decision that is up to parents. Additionally, this measure relies on a simple “yes” or “no” response and therefore should not be biased by national differences in response styles.

This section has reviewed the evidence on the East Asian-U.S. parental involvement gap and has argued that while the evidence given to support the idea that U.S. parents are less effectively involved in their children’s education than parents in East Asian nations is weak, there is some compelling evidence to suggest that a greater proportion of parents in Japan and Taiwan than in the U.S. is involved in their children’s education. Therefore, I use the phrase “East-Asian-U.S. parental involvement gap” to refer solely to a cross-regional disparity in the proportion of involved parents. In the next section, I review the prevailing theory to explain why a larger proportion of parents in East Asian nations than in the U.S. is involved in their children’s education before showing how the theory can be improved.

PREVAILING THEORY ON THE EAST ASIAN-U.S. PARENTAL INVOLVEMENT GAP

The prevailing theory on the East Asian-U.S. parental involvement gap is supported by both cross-regional and immigrant studies (for example: Chen and Stevenson 1995; Kao 1995; Kao and Thompson 2003; Okagaki and Frensch 1998; Schneider and Lee 1990; Stevenson and Stigler 1992; Stevenson, Chen and Lee 1993; Sun 1998). According to these studies, the reason why a greater proportion of East Asian parents than U.S. parents is involved in their children’s education is because East Asian culture prompts parents to regard doing well in school as their children’s most pressing task (Stevenson and Stigler 1992). Scholars note than in Confucianism, education is important for family honor, self-esteem and self-improvement (Schneider and Lee

1990). These strong cultural beliefs concerning the value of education supposedly drive East Asian parents to become involved in their children's education.[27]

The prevailing argument is that how much parents value education is rooted in their national culture and determines whether or not they become involved in their children's education. Given that immigrant studies of parental involvement also tend to attribute the greater educational investments of immigrant East Asian parents compared to other racial/ethnic groups to East Asian culture, the implication is that Confucianism is unique among world cultures in the emphasis it places on education. This literature yields the first hypothesis to be examined in the analysis section:

> *Hypothesis 1*: The proportion of parents who are involved in their children's education is uniquely high in East Asia, given that it is the region where Confucianism is most prevalent.

There are significant weaknesses to the prevailing theory. For one, it rests on three *a priori* assumptions that have little if any empirical foundation: 1) Parents' beliefs about the value of education determine whether or not they become involved in their children's education, 2) Beliefs about the value of education are rooted in national cultures, and 3) U.S. culture places less value on education than East Asian culture. As for the first assumption, U.S.-based research on parental involvement has looked for but cannot find evidence to support the idea that whether or not parents value education determines their degree of involvement (Chin and Phillips 2004; Hoover-Dempsey et al 2005; Lareau 2000[1989]). The second assumption suggests that beliefs about the importance of education are inscribed by "nature" as opposed to "nurture," or by place of birth as opposed to structural or institutional factors. It is a defeatist argument; the implication

[27] A passage from Lin and Fu (1990: 429) demonstrates the degree to which scholars attribute Asian parenting practices to Confucianism: "Definitive views on parental control, obedience, strict discipline, emphasis on education, filial piety, respect for elders, family obligations, reverence for tradition, maintenance of harmony, and negation of conflict are attributed to the influence of Confucianism."

is that the proportion of parents who are involved in their children's education will always be higher in East Asian nations than in the U.S.

The third assumption—that U.S. culture places less value on education than East Asian culture—indicates that the prevailing theory is based on a cultural depravity model. This type of model was used to explain the lower performance of African-Americans in the U.S. public school system (e.g. Moynihan 1965) but fell out of fashion in sociology in the 1970s (Corcoran 1995). Yet alternative explanations for the parental involvement gap between parents in East Asia and the U.S. are virtually nonexistent (Kao and Thompson 2003).

Another limitation to the prevailing theory is that it does not allow for factors other than how much a national culture values education to influence whether or not parents become involved in their children's education. For example, cross-regional research rarely considers how schools might influences parents' decisions to become involved in their children's education. Stevenson and Stigler's (1992) work is a prominent exception. The authors note that elementary schools in both Japan and Taiwan engage parents in the schooling process through the daily use of a communications booklet, which updates parents on their children's progress and keeps them abreast of what is going on at school. The authors acknowledge that this school practice facilitates the involvement of Japanese and Taiwanese parents, writing (p. 84), "Only through this degree of intimate knowledge of what is happening at school can parents hope to be aware of the ways they can help their children and provide a home environment that is conducive to studying." Yet the authors ultimately attribute the greater proportion of involved parents in Japan and Taiwan to Confucianism as opposed to the communications booklets. In other words, school programs meant to engage parents in the schooling process in Taiwan and Japan do not actually encourage parents to become involved in their children's education because, thanks to

Confucianism, all parents in East Asia are already involved. The cumulative effect of these weaknesses in the prevailing theory is that scholars too readily conclude that nothing can be done to increase the proportion of parents who are involved in their children's education in the U.S. In the next section I explain, using sociological theories from U.S.-based research, what can be done.

AN ALTERNATE THEORY ON THE EAST ASIAN-U.S. PARENTAL INVOLVEMENT GAP

Case studies of preschools and elementary schools in Japan published since Stevenson and Stigler's (1992) book provide the empirical groundwork for a more positive assessment of what schools and educational systems can do to increase the proportion of parents who are involved in their children's education. These studies do not rest on the assumption that Confucianism propels all parents of East Asian descent to become involved in their children's education. Instead, they provide reason to believe that the communications booklets and other widely-used school practices to engage parents in the schooling process in Japan are succeeding in encouraging parents, who would otherwise take a backseat to their children's education, to become involved. For example, to analyze variation in parental involvement across social-class backgrounds in Japan, Yamamoto, Holloway and Suzuki (2006) interviewed 108 mothers of 5- and 6-year-olds in two urban centers. They find that while highly educated mothers tend to use a more rigorous preschool selection process and to spend more time reading to their children at home than their less-educated counterparts, the group of mothers with less education tends to spend more time participating in school-based activities. The reason why the group of less educated mothers actually spend more time at school, the authors suggest, is that all mothers, but particularly those with less education, find it difficult to say no to requests from school staff to participate or volunteer at school. The authors' research raises the possibility that the invitations to involvement from school staff are increasing the proportion of mothers who are involved in their children's education in Japan.

Additional support for the idea that widespread, school-based practices to engage parents in the schooling process in Japan and Taiwan are encouraging parents who might otherwise take a backseat to their children's education to become involved comes from four anthropological studies of preschools and elementary schools in Japan written by Americans (Allison 1996, Benjamin 1997, Lewis 1995, Peak 1991). Three of the four authors (Allison, Benjamin and Lewis) enrolled their own children in school in Japan and provide firsthand accounts of their interactions with their children's schools.[28] All four studies provide similar descriptions of the formal mechanisms through which schools in Japan engage parents in the schooling process. These mechanisms are not limited to the communications booklets that Stevenson and Stigler (1992) note, but also include frequent face-to-face meetings, invitations to volunteer at school, and other written directives from school to home. It is possible, although unclear from these studies, that many of these mechanisms are institutionalized at the system level. For example, the educational system in Japan requires parents to meet with the school principal before their child enters a new school.[29] Regardless, all four studies suggest that compared to elementary schools in the U.S., elementary schools in Japan go to greater lengths to encourage parental involvement.

Two of the authors explicitly attribute the high proportion of involved parents in Japan to schools' efforts to engage parents in their children's education. Benjamin (1997) found the mass of written communication from school to be particularly effective in enlisting her support,

[28] I looked for, but could not find, similar accounts of preschools and elementary schools in Hong Kong, Singapore and Taiwan.

[29] Information last retrieved on March 12, 2012 from www.education-in-japan.info/sub106.html

writing (p. 196), “the volume of information and advice and demands [from the school] resulted in a higher level of cooperation. Maybe every mother could decide to ignore 20 percent of what the school suggested or demanded, but following 80 percent of a hundred directions means that mothers are doing a lot.” From Allison’s (1996) perspective, the demands of Japanese schools on mothers are so extensive—and so effective—that she titled her piece “Producing Mothers.” The aim of Allison’s work is to combat the prevailing theory that Japanese mothers decide on their own to become involved in their children’s education. Instead, her central thesis is that the Japanese school system coerces mothers to play a leading role in their children’s education. Whereas the prevailing theory on the East Asian-U.S. parental involvement gap attributes the greater proportion of involved parents in East Asian nations than in the U.S. to Confucianism, which acts like an internal rudder guiding East Asian parents to act for the benefit of their children’s education, Allison attributes the high proportion of involved parents in Japan to the Japanese educational system, which does its best not to let parents (particularly mothers) take a backseat to their children’s education.

Case studies of Japanese preschools and elementary schools suggest the possibility that the East Asian-U.S. parental involvement gap stems from the relative inability of schools within the U.S. to foster parental involvement. If this is true, then within each nation we would expect to find that the likelihood of involvement is greater at schools with parental involvement programs than at those without such programs, and that parental involvement programs are relatively less extensive or less successful in the U.S. The lack of data on the specific strategies schools use to encourage parental involvement makes it impossible to directly assess the impact of various school-based programs on the likelihood of parental involvement within nations, and to compare the scope of parental involvement programs across nations.[30] Nonetheless, it is possible to assess whether or not the likelihood that parents from similar backgrounds are involved differs across schools within each nation. If so, then this chapter would provide some support for the alternate argument. Therefore, the second hypothesis this chapter examines is:

> *Hypothesis 2*: Controlling for parents' background, the likelihood that parents are involved in their children's education varies across schools within each nation.

While the case studies leave open the possibility that schools can make a difference in the proportion of parents who are involved in their children's education, they do not explain why schools might make a difference. For this reason, I use Coleman (1991) and Epstein's (1986) work to put research of the East Asian-U.S. parental involvement gap on sounder theoretical footing.

Both Coleman (1991) and Epstein (1986, 1995) draw connections between the social capital in the school community and the proportion of parents who are involved in their children's education. As such, a brief description of the concept of social capital is in order. Coleman (1988) conceived of social capital as taking multiple forms, all of which have in common two features: 1) they consist of a social organization or network, and 2) they facilitate certain actions of actors within the social organization. The three forms of social capital that Coleman elaborated on were obligations and expectations (e.g. mutual trust among actors in the social organization), information channels, and norms that encourage action for the common good. Coleman contends that social closure, or strong social ties between all actors in the network, facilitates the creation of trust and of social norms that lead to positive social control. Coleman's concept of social capital draws attention to the role that social structure plays in

[30] Initially I planned to use school principals' responses to questions regarding the activities schools ask parents to take part in to differentiate schools based on the strength of their parental involvement programs. Unfortunately, these data proved inadequate as principals' responses do not vary within nations.

determining access to resources, and the power that social structure has over individuals' actions.[31]

Scholars have pointed out that Coleman did not distinguish between the resources individuals gain through the social structure and the social structure itself, which leads to tautological arguments (Portes 1998). A case in point is that Coleman referred to parental involvement as both social capital and as an outcome of social capital. For this reason, I find it useful to keep the concept of parental involvement distinct from the concept of social capital and to decompose the latter into three components: 1) resources that accrue to individuals ("individual resources"), 2) resources that stay in the social network and are available to all actors within the social organization ("social resources"), and 3) the social organization itself.[32] Examples of resources that individuals might gain through a social network are job offers, marriage partners, and as Coleman emphasizes for children, human capital. Examples of what I call social resources are mutual trust and respect, information channels, and high achievement norms. Decomposing the concept of social capital into three components makes it possible to discuss the importance of each aspect of social capital with greater clarity, and helps to distinguish between the causes and consequences of social capital.

Coleman (1991) pointed to intergenerational social closure and information sharing as the forms of social capital, or social resources, that most influence the proportion of parents who are involved in their children's education at each school. Coleman (1988; 1991) believed that

[31] Dika and Singh (2002) rightly point out that Coleman's conceptualization of social capital highlights the importance of social norms and positive social control. However, it is clear that Coleman also viewed social capital as a means of accessing resources. The main point of his 1988 article "Social Capital in the Creation of Human Capital," is that without "social capital in the family," (e.g. the time and effort parents spend with children on intellectual matters), children may not have access to their parents' human capital.

[32] Bourdieu (2000[1986]) provided a contemporaneous conceptualization of social capital which emphasizes the importance of social capital in accessing scarce resources. Each person's social capital is defined by the number of people in his social network and the amount of economic and cultural capital possessed by each person in the network. While this conceptualization provides greater clarity on how to measure social capital, it neglects the importance of the volume of *social resources* within the network.

intergenerational closure—a dense social organization characterized by strong ties among parents in the school community—leads to the creation of shared norms and standards regarding children's behavior, imposes strong social sanctions on deviant behavior, and incentivizes parents and children to work for the common good of the school community. In school communities characterized by intergenerational closure, informal social sanctions against nonparticipation in school-related activities compel parents—even those who prefer to take a backseat to their children's education—to become involved. Coleman urged schools that lack intergenerational closure (or strong PTA associations) to create it by bringing parents in the school community together on matters related to their common interests (such as having children in the same grade).

The other form of social capital that Coleman (1991) identified as most likely to influence the proportion of involved parents in the school community is information sharing. Apparently, Coleman perceived of parents' decisions to become involved in their children's education as depending in large part on their knowledge of what their children need to succeed in school and of how best to support their children's education. Coleman (1991: 18) explained, "parents are unskilled in helping their children to succeed in school. Even well-educated parents often lack the knowledge of what practices in the home will most help their children to succeed in school." By sharing this information with parents, schools can help parents become aware of the importance of their involvement in their children's education and in so doing encourage parents who would otherwise take a backseat to their children's education to become involved.

The quantitative data available at the time Coleman developed his theory did not allow him to satisfactorily test the relationship between social capital in school communities and parental involvement. The best evidence to support his hypotheses comes from his research on

Catholic schools (Coleman and Hoffer 1987: 55), which shows that levels of parental involvement, as measured by attendance at PTA meetings, parent-teacher conferences and volunteering for school projects, are higher on average at Catholic schools than at public schools, even after family socioeconomic status (SES) is taken into account.

Empirical assessments of Coleman's theory since help to refine his argument and highlight the data still needed to adequately test his theory. For example, Morgan and Sørensen's (1999) analysis of the NELS data suggest that Coleman overemphasized the importance of intergenerational closure for parental involvement. Their findings indicate that intergenerational closure, measured at the school level as the average number of parents of children's friends from school that parents know, does not entirely explain the Catholic school advantage in parental involvement and is negatively associated with student achievement in public schools. The authors challenge Coleman's idea that intergenerational closure automatically incentives parents to act to benefit their children's education. The implication is that the level of intergenerational closure in the school community, on its own, does not predict the proportion of involved parents at school.

Scholars have also criticized Coleman for overlooking the possibility that non-dominant groups are excluded from the social structure of the school community (Dika and Singh 2002; Stanton-Salazar 1997, with Dornbusch 1995). Parents who are excluded from the social structure are denied access to the social resources that are theorized to influence parental involvement. This is an important point because it suggests that even if social resources adhere in the structure of relations in the school community, the proportion of parents who are involved in their children's education might still be low because a large portion have been excluded from the social structure. These criticisms suggest the need for empirical evidence not only of whether or not schools encourage parents to interact with one another but also of whether the programs target all parents or only a fraction of them.

Similarly, some research suggests that there is considerable variation in the effectiveness with which schools might share information with parents and create opportunities for parents to interact with one another. Hoover-Dempsey (et al 2005), whose primary research focuses on the factors motivating parents to become involved in their children's education, reports that it is difficult for teachers to encourage effective involvement—particularly when the backgrounds of the parents they are trying to reach differ dramatically from one another. This is because a meaningful invitation to involvement needs to target *all* parents and to include strategies for involvement that might be implemented by parents from a range of educational backgrounds. An ill-prepared invitation to involvement could have the unintended consequence of excluding some parents from becoming involved in their children's education. The implication is that the correspondence between parental involvement and information sharing might be weaker than Coleman anticipated. Scholars hoping to test Coleman's hypothesis that information sharing is associated with higher proportions of involved parents might require more fine-grained data on the tactics schools use to share information with parents as well as on the content of the information that is shared.

Lareau's (2001[1989]) ethnographic study of two elementary schools in California also raises the possibility that the proportion of involved parents in the school community might not relate to the extent of information sharing in the school community. This is because parents might gain access to the knowledge they need to help their children succeed in school from their personal social networks. Lareau reports that the upper-middle class parents in her study most often relied on their personal social networks to learn how to best support their children's education. On the other hand, the working class parents in her study were more apt to lack social ties to knowledgeable others and thus tended to rely almost exclusively on the school for this information. Lareau's findings support Coleman's notion of the importance of knowledge for parental involvement but imply that the need for schools to share information in order to increase the proportion of involved parents is less pressing at high SES schools than at low SES schools.

To sum, according to Coleman (1991), schools can make a difference in the proportion of parents who are involved in their children's education by sharing information about successful forms of parental involvement with parents and by helping to build close ties among parents (which in turn should incentive parents to be involved in their children's education). Research published since calls into question the simiplicity of the connections Coleman draws between social resources and the proportion of involved parents in the school community and suggests that empirical tests of Coleman's theory would require more data than are currently available. Fortunately, Epstein's (1986, 1995) theory offers some hypotheses that we can begin to assess using the TIMSS data.

As the Director of the Center on School, Family, and Community Partnerships at Johns Hopkins University and the founder and Principal Research Scientist at the National Network of Partnership Schools, Epstein has dedicated her professional career to the research and development of programs to increase the proportion of parents who are effectively involved in their children's education. Similar to Coleman (1988, 1991), Epstein's research (1986, 1995) underscores the importance of social resources in school community to promote parental involvement. However, Epstein draws attention to another social resource in the school community—mutual trust and respect—that she argues is crucial for school administrators to develop if they hope to increase the proportion of involved parents at their school. Additionally, unlike Coleman who emphasized social ties among parents, Epstein focuses on family-school relationships.

According to Epstein (1986), family-school relationships must be grounded in mutual trust and respect if they are to increase the proportion of parents who are involved in their children's education. Epstein bases this idea on a simple premise of reference group theory, which claims that individuals and groups take into account the actions and ideas of those they know and recognize as important. Therefore, to influence parents' actions and behaviors towards their children's education, school personnel must earn the trust and respect of all parents. Similarly, because trust and respect needs to be mutual between parents and school staff, staff must view parents as their partners in children's education and development. Epstein's emphasis on family-school relationships for encouraging parents to become involved in their children's education leads to the third hypothesis that this chapter will empirically assess:

> *Hypothesis 3*: The quality of family-school relationships in the school community influences the likelihood that parents become involved in their children's education, net of parents' background.

Given the importance of mutual trust and respect in the school community for parental involvement, it is possible that the level of concentrated disadvantage in a school community also affects the likelihood of involvement. This is because concentrated disadvantage in a community has been shown to stymie the development of mutual trust and respect or "collective efficacy." Using data from the Project on Human Development in Chicago Neighborhoods 1995 from over 8,000 residents representing 343 neighborhood communities in Chicago, Sampson, Raudenbush and Earls (1997) reveal a strong and statistically significant, negative association between concentrated disadvantage and collective efficacy that cannot be attributed to the aggregated demographic characteristics of individuals. This argument provides the fourth hypothesis that I will test in this chapter:

> *Hypothesis 4*: Concentrated disadvantage in the school community decreases the likelihood that parents are involved in their children's education, net of parents' background.

The alternative argument suggests that two features of the school community explain variations in the likelihood that parents are involved in their children's education that are not accounted for by parent background factors. Specifically, this argument suggests that high-quality family-school relationships increase and concentrated disadvantage decreases the likelihood of involvement in children's education. The question left to ask is why schools in the U.S. seem to lag behind those in Japan and Taiwan when it comes to encouraging parents to become involved in their children's education.

One reason might be that there are more school communities characterized by concentrated disadvantage in the U.S. than in East Asian nations. This would make it more difficult, on average, for schools in the U.S. to foster family-school relationships that are based on mutual trust and respect. To assess the merit of this possibility, I empirically analyze the following hypothesis:

> *Hypothesis 5*: There are fewer school communities characterized by concentrated disadvantage in East Asian nations than in the U.S.

Another reason why schools in the U.S. might be less adept at fostering parental involvement than schools in Japan or Taiwan might be because educators in the U.S. have not receiving adequate training in how to encourage parents to become involved in their children's education. Epstein (1995) reports a dearth of college and university courses geared towards providing future educators with an understanding of the importance of partnerships with parents, and the tools they need to develop, evaluate, guide and implement successful parental involvement programs. Similar studies of teacher-training programs in the East Asian nations under investigations are not readily available. However, given the reported institutionalized nature of parental involvement programs in Japan and Taiwan (Allison 1996, Benjamin 1997, Lewis 1995, Stevenson and Stigler 1992), it seems highly likely that educators in these nations have been made aware of the importance of involving parents for their occupational success.

Both the lack of professional training and of institutions such as the communications booklets of Japan and Taiwan could mean that school staff in the U.S. do not perceive the job of teaching parents how to help their children succeed in school as part of their occupational responsibilities. Indeed, Lareau's (2000[1989]) ethnographic study of two first-grade classrooms, one at a predominantly white middle-class public school and the other at a predominantly white working-class public school, provides evidence to suggest that school staff in the U.S. view what parents do to support their children's education as something over which they have no control. Lareau (2000[1989]) reports that personnel at both schools interpret parental involvement as a sign of how much parents value education, and tend to associate parental involvement with social class. As the principal at the working-class school explained to Lareau (2000: 98), parents who are not involved in their children's education, "don't value education because they don't have much of it themselves." Lareau also notes that teachers at the middle-class school attribute the ease of their face-to-face meetings with parents to parents' high social status.

If school staff in the U.S. typically view parents' decisions to become involved in their children's education as determined by social class, then it is unlikely that they perceive of parental involvement as something that they can influence. In turn, they likely devote less of their time and energy to encouraging parents who are not already involved in their children's education to become involved. This could explain why the proportion of involved parents is lower in the U.S. than in Japan and Taiwan if school staff in those nations do not share the same perceptions (e.g. if staff in Japan and Taiwan do not link parental involvement to social class background). This argument leads to the sixth and final hypothesis that this chapter will examine:

> *Hypothesis 6*: Staff perceptions of parental involvement are influenced by parents' social class background more so in the U.S. than in East Asian nations.

In the following sections I briefly outline the data and variables before moving on to the analyses. The analyses cover a range of hypotheses regarding relationships at the national, school and individual level. Tests of the first hypothesis examine the proportion of parents who are involved in their children's education at the national level. Findings from this analysis let us know whether the disparity in the proportion of parents who provide their children with a desk at home continues to be higher in Japan and Taiwan than in the U.S., and where the percentages of involved parents in East Asian nations rank among all high-achieving nations that participated in either the 2003 or 2007 TIMSS fourth-grade population. Including other high-achieving nations in tests of Hypothesis 1 provides the opportunity to begin to address a key tenet of the prevailing theory on the East Asian-U.S. parental involvement gap, which is that national cultural beliefs determine the proportion of parents who are involved in their children's education.

Next, tests of the second, third, and fourth hypothesis focus on the relationship between features of the school community and the likelihood that individual parents are involved in their children's education. By conducting these analyses separately for the U.S. and the four East Asian nations under investigation, I am able to shed light on the universality of the alternative argument I presented to explain the East Asian-U.S. parental involvement gap. Lastly, tests of the fifth and sixth hypothesis are meant to provide insight into the reasons for the East Asian-U.S. parental involvement gap. These analyses focus on cross-regional variations in the features of the school community that are expected to influence involvement.

DATA

To put research of the East Asian-U.S. parental involvement gap into comparative perspective, I use data from all high-achieving nations that participated in either the 2003 or 2007 round of the TIMSS fourth-grade population. I define a high-achieving nation as one with an average score on the TIMSS math assessment that is at or above the international mean of 500. The nations included in this analysis are: Australia, Austria, Belgium, Denmark, England, Germany, Hong Kong, Hungary, Italy, Japan, Latvia, Lithuania, the Republic of Moldova, the Netherlands, Singapore, Sweden, Taiwan and the U.S. Cases with missing data that could not be obtained from another source were dropped from analysis.[33]

VARIABLES

I have argued that one of the most compelling pieces of evidence that an East Asian-U.S. parental involvement gap exists is that a greater proportion of parents in Japan and Taiwan than in the U.S. had provided their fifth graders with a desk at home (Stevenson and Stigler 1992). I base the empirical analyses of this chapter on the same measure of *parental involvement*. Specifically, the TIMSS Student Questionnaire asks students to respond “yes” or “no” to the question of whether they have a study desk or table to use at home.

[33] Many of the variables in this study rely on information that was collected from multiple sources within a school. As long as the information was provided by one source within a school, the cases were not dropped from analysis.

To confirm that provision of a study desk acts as a respectable indicator of parental involvement for each East Asian nation and the U.S., I compare the mean score of the group of students with a desk at home to the group without on measures of students cognitive abilities, non-cognitive attributes, and additional indicators of parental involvement. I use adjusted Wald tests of significance to ascertain whether or not the difference in means between the two groups of students is statistically significant. To ensure that the relationship between provision of a study desk at home and student outcomes cannot be attributed to family SES, I also calculated within each family SES stratum the difference in means between students with a desk and students without. (As in the Teacher chapter, SES is measured using information on the number of books students have at home.) Table 4.2 displays the results of these tests for the full sample, the sample of low SES students, and the sample of average SES students. Results for the whole sample are provided in the first two columns of data, results for the sample of low SES families are given in the next two columns, and results from the sample of average SES families are provided in the last two columns. Results for each nation are provided in a separate panel.

According to Table 4.2, within each nation provision of a desk is positively associated with students' cognitive abilities in all nations except Hong Kong. In the U.S., Japan, Singapore and Taiwan the group of students with desks have a higher average math achievement score than the group of students without desks. In the whole sample for these nations, the difference in mean achievement between students with a desk and those without is statistically significant (at the $p<0.001$ level). The difference in mean achievement between the groups of students with and without a study desk at home is statistically significant (at the $p<0.05$ level or smaller) in the low SES and average SES samples in each of these nations as well. For example, among students of low SES background in Singapore, the group with desks at home averages a score of 543 on the

TIMSS math assessment, while the group without desks averaged a score of 504. The difference of 39 achievement points is statistically significant (at the p<0.001 level).

Table 4.2 Mean Characteristics of Students With and Without Desks at Home

	Whole Sample		Low SES		Average SES	
Nations	**Without desk**	**With desk**	**Without desk**	**With desk**	**Without desk**	**With desk**
United States						
Achievement	498	532***	470	484***	506	534***
Enjoy Math	3.12	3.22***	3.13	3.17	3.13	3.23***
Like School	3.02	3.18***	3.00	3.05	3.05	3.19***
Watch TV	3.18	3.11**	3.21	3.30	3.16	3.09**
Read Books	2.25	2.44***	2.04	2.12	2.30	2.40***
Hong Kong						
Achievement	593	589	577	567*	599	592**
Enjoy Math	2.94	2.99*	2.82	2.83	2.98	3.01
Like School	2.82	2.92***	2.73	2.74	2.85	2.95**
Watch TV	3.14	3.03***	3.26	3.22	3.11	3.01**
Read Books	2.30	2.49***	2.03	2.07	2.36	2.51***
Japan						
Achievement	549	568***	507	528*	555	567**
Enjoy Math	2.83	2.90	2.77	2.75	2.80	2.90
Like School	3.10	3.18*	3.01	3.01	3.13	3.19
Watch TV	3.35	3.22**	3.34	3.32	3.36	3.23*
Read Books	2.09	2.17	1.86	1.80	2.10	2.13
Singapore						
Achievement	558	602***	504	543***	575	603***
Enjoy Math	3.13	3.31***	3.09	3.22	3.14	3.32***
Like School	3.38	3.46*	3.33	3.41	3.42	3.47
Watch TV	3.36	3.22***	3.45	3.34	3.36	3.23**
Read Books	2.37	2.61***	2.05	2.14	2.44	2.56*
Taiwan						
Achievement	555	573***	525	534*	567	575*
Enjoy Math	2.61	2.78***	2.53	2.66	2.63	2.79***
Like School	2.90	3.03***	2.88	2.95	2.89	3.03***
Watch TV	2.67	2.54***	2.73	2.70	2.69	2.55**
Read Books	2.25	2.47***	2.07	2.11	2.30	2.44***

*** p<0.001, ** p<0.01, * p<0.05; Placed after the mean value for students with desks at home if adjusted Wald tests of significance indicate that the difference in means between the group of students with desks is statistically significant different from the mean of the group of students without desks.

In Hong Kong, the relationship between having a desk and achievement is reversed. The difference in average achievement between the two groups of students is small (4 points in the whole sample, 10 points in the low SES sample and 7 points in the average SES sample) but statistically significant for the low and average SES subsamples.

Table 4.2 also reports the mean score on measures of non-cognitive attributes for the groups of students with and without a study desk at home. Within each nation, responses to the statements "I enjoy learning mathematics" and "I like being in school," which could range from "disagree a lot" (score=1) to "agree a lot" (score=4) are higher (e.g. more positive) for the group of students with desks than for the group without. In the whole sample for each nation, the difference in average responses between the two groups of students is statistically significant (at the $p<0.05$ level or smaller) in all nations save Japan (where the difference is statistically significant for liking school but not for enjoying math). The differences are not statistically significantly different in any of the low SES samples, but some are statistically significant in the average SES samples. In the U.S. and Taiwan, the difference between the group of average SES students with a desk and those without a desk is statistically significant (at the $p<0.001$) for both enjoying math and liking school; In the average SES sample in Singapore the difference in enjoying math is statistically significant (at the $p<0.001$ level), and in the average SES sample in Hong Kong the difference in liking school is statistically significant (at the $p<0.01$ level). These findings indicate a positive association between provision of a study desk and non-cognitive attributes that have been linked to success in school (Rothstein 2004).[34]

Scholars have sometimes relied on the amount of time that children spend reading books for enjoyment or watching TV outside of school as indicators of more or less, respectively,

[34] Note that the findings do *not* indicate a causal relationship between provision of a desk and non-cognitive attributes linked to academic success. It could be that parents are more likely to provide a desk to children who display a taste for studying.

parental involvement (Fan and Chen 2001). Table 4.2 shows that in most cases, the group of students with a desk at home typically spent more time reading books for enjoyment and less time watching TV per week than the group of students without a desk.[35] In the average SES and whole samples, the differences between groups are statistically significant (at the $p<0.05$ level or smaller) in all nations except Japan, where the difference is statistically significant only for TV watching. These findings indicate relationships in the appropriate direction between provision of a desk and other measures of parental involvement, which lends additional support to the idea that provision of a study desk acts as a respectable measure of parental involvement. Although provision of a desk is not related to higher math achievement in Hong Kong, it is positively associated with non-cognitive attributes, and with other indicators of parental involvement. Therefore, familial provision of a desk acts as a fair measure of parental involvement in Hong Kong as well as in the other nations under investigation.

To measure the proportion of parents who are involved in their children's education at each school, I divide the number of students who have a study desk at home by the total number of students surveyed at each school. For a nation-level measure of the proportion of involved parents, I calculate the percentage of fourth-grade students in each nation who responded that they had a study desk or table to use at home. I perform these calculations using population weights that take into account the stratified nature of the TIMSS sampling design to provide measures of the proportion of involved parents that are nationally representative.

I rely on two questions posed to school principals regarding the relative percentages of students in their schools who come from economically advantaged and disadvantaged homes to measure the degree of *Disadvantage* in the school community. Responses to both questions were

[35] The only instance in which the relationship is reversed is in the low SES sample from Japan. Here, the group of students without desks average more time reading books for enjoyment than the group of students with desks, although the difference in means is not statistically significant.

limited to four categories: "0-10%," "11-25%," "26-50%," or "more than 50%." I recode each variable with values equal to the midrange (e.g. responses of "0-10%" are given a value of 5) and code responses of "more than 50%" as 63. Next, I subtract the percentage of economically advantaged students from the percentage of economically disadvantaged students at each school. Thus, this measure differs from typical measures of school SES in that higher scores indicate concentrated disadvantage. For example, the highest score on this variable is 58, which means that more than 50 percent of the student population comes from economically disadvantaged homes and 0-10 percent come from economically advantaged homes. The low score of -58, meanwhile, indicates a high proportion of advantaged students relative to disadvantaged students at the school. Finally, I create two dummy variables to indicate schools that are one standard deviation above or below the mean level of disadvantage.

Criminal or violent activity tends to be higher in communities of concentrated disadvantage (Sampson et al 1997). Therefore, I create another indicator of the degree of disadvantage in the school community using teachers' responses to the following statements: 1) "This school is located in a safe neighborhood" and, 2) "I feel safe at this school," which could range on a 4-point scale from "Agree a lot" to "Disagree a lot." The Cronbach's alphas for this scale measure of *Neighborhood Violence* range from 0.65 for the Japan data to 0.86 for the Singapore data. On this scale, higher values indicate higher-than-average levels of violence or crime in the school neighborhood.

To gauge the quality of parent-school relationships at each school, I rely on two questions that are included in both the School Principal Questionnaire and the Teacher Questionnaire. These questions ask respondents to characterize the level of "Parental support for student achievement" and "Parental involvement in school activities" at their school on a 5-point scale ranging from “very low” to “very high.”[36] I combine teacher and principal responses to make this a school-level variable.[37] Cronbach’s alphas for this measure range from 0.76 in Taiwan to 0.91 in the U.S., suggesting that the two questions on which this scale is based are tapping into the same latent construct.

It is also possible that language barriers within the school community restrict the depth and strength of the social relationships that can be developed among parents and between parents and school staff. To control for this possibility, I create a proxy measure of the *Language Barriers* within each school community based on school principals’ rough categorization of the percentage of students enrolled in school who do not speak the language of the TIMSS assessment at home. Responses could be, “More than 90%,” “75 to 90%,” “50 to 75%,” or “Less than 50%.” Higher values on this scale represent schools with more linguistic barriers for school community to overcome. In the case of Singapore, this variable is not indicative of the language barriers in the school community because the TIMSS assessment was given in English. While English is the first language of the nation, Singapore has three other official languages (Chinese, Malay and Tamil) that are often the main language spoken in homes, and students in Singapore generally attend schools that offer instruction in both English and their mother-tongue. (For example, ethnic Chinese in Singapore typically attend public schools that offer instruction in Chinese.) For the Singaporean analyses I use the percentage of students in the TIMSS sample whose parents are immigrants as this provides some insight on, perhaps, the cultural barriers

[36] These are similar to the survey questions that Lee and Burkan (2003) used to measure the positive relationships between teachers and students.

[37] Research on social organizations offers a more accurate and reliable method for aggregating responses, which takes into account measurement error due to item and rater inconsistencies (Raudenbush, Rowan and Kang 1991; Raudenbush and Sampson 1999a, 1999b). However, this method requires data from more than 5 respondents from each organization (Harding 2011). As there are only, on average, two teachers surveyed per school in each nation, this method is not a feasible option for these data.

parents and school staff must overcome to work together to advance children's educational development.

At the family level, I control for students' age and sex, as well as the language the family speaks at home, immigrant status, and family SES. The variable for family SES indicates whether or not the number of books students have at home falls one standard deviation above or below the national mean. The variable for immigrant status indicates that one or both of students' parents were born outside the nation and/or that the student did not immigrate to the nation before the age of one. A score of 1 on the non-native language variable indicates that families rarely speak the language of the test at home, while a score of zero indicates that students always or almost always speak the language of the test at home. For all nations but Singapore, this variable indicates that families are non-native language speakers. For Singapore, this variable provides insight into whether or not families prefer to speak with one another in English at home.

At the school level, I include two dummy variables indicating the size of the school community. The first indicates that schools are located in *Large Communities* of more than 100,001 people. The second indicates schools in *Small Communities* of fewer than 15,000 people. This variable is not included in analyses of the Singapore data because all school principals' responses to the question of school community were gang punched to the same answer. I also include one variable indicating the *School Size* (e.g. the total number of students enrolled in the school). Lastly, to control for the sample round I include a dummy variable indicating whether or not data came from the 2003 TIMSS (as opposed to the 2007 round).

RESULTS

Are the proportions of involved parents uniquely high in East Asia?

Stevenson and Stigler (1992) demonstrate the East Asian-U.S. parental involvement gap by showing that the proportion of parents who had provided their fifth grader with a desk at home is higher in Japan and Taiwan than in the U.S. Beyond that, however, international comparative research has yet to show that the proportion of involved parents is uniquely high in East Asia. This is important because the prevailing theory on the East Asian-U.S. parental involvement gap suggests that national cultural beliefs are the principal motivating factor of parental involvement and that Confucianism is exceptional among world cultures in the premium it places on education. If this were true, then the proportion of involved parents would be higher in East Asian nations than in other nations. To test this theory, I calculate the percentage of students with a desk at home in each high-achieving nation and display them in Figure 4.1. The results confirm Stevenson and Stigler's (1992) finding that a greater proportion of parents in Japan and Taiwan than in the U.S. provides their children with a desk at home, but casts doubt on the argument that Confucian culture sets East Asian parents apart.

According to Figure 4.1, the percentages of parents who provide their fourth grader with a desk at home are not abnormally high in East Asian nations compared to other high-achieving nations. In fact, the percentages are greater in Austria (98.1%), Germany (98.0%), Sweden (97.3%), Lithuania (97.2%), and Hungary (95.6%) than in any East Asian nation. Further, the percentage of parents who provide their children with a desk at home varies considerably within East Asia. Among East Asian nations, Japan boasts the highest percentage of students with a study desk (94.0%), while the percentage is around the international mean of 88.1 percent in Singapore (89.2%) and Taiwan (86.7%) but considerably lower in Hong Kong (71.5%). In fact, the percentage of parents who provide a desk at home is lower in Hong Kong than in the U.S. (77.4%).

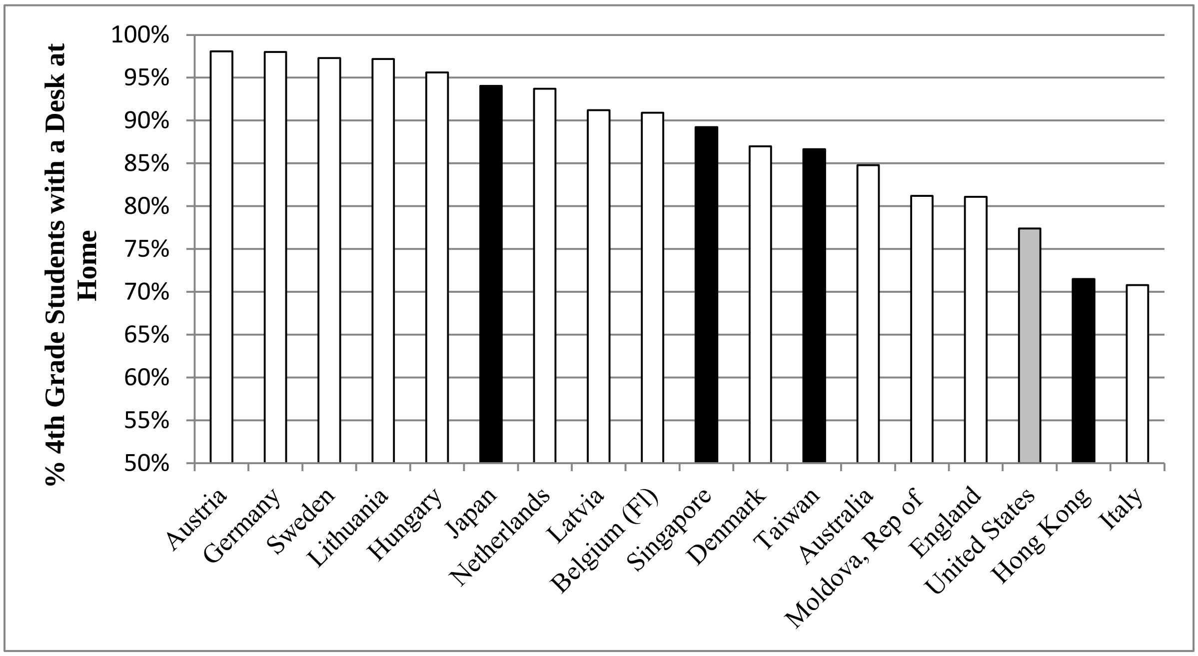

Figure 4.1 Percentage in each High-Achieving Nation with a Desk at Home

The differences across East Asian nations and the U.S. in the proportion of parents who provide a desk at home are not simple reflections of cross-national differences in income inequality. If the proportion of parents who provide a desk at home were to mirror the level of income inequality in each nation, then the Gini coefficient would be highest in Hong Kong followed by the U.S., Taiwan, Singapore and Japan. This is not the case. According to Deininger and Squire (1996), the Gini coefficient of income inequality over the latter half of the twentieth century is roughly equivalent between the U.S. (35.3) and Japan (34.8), highest in Hong Kong (41.6) and Singapore (40.1) and lowest in Taiwan (29.6).

Overall, the results indicate that among the higher achieving nations that participated in the TIMSS fourth-grade population, parental involvement, as measured by the provision of a desk at home, is lowest in Italy, Hong Kong and the U.S. and highest in Europe. The percentage of parents who are uninvolved, according to this measure of parental involvement, is about 23 percent in the U.S., which is roughly four times the percentage of uninvolved parents in Japan and twice that of Singapore and Taiwan. The implication of these findings is that Confucianism is not a prerequisite to obtaining high levels of parental involvement within a nation.

Do schools influence the likelihood of involvement?

The alternative explanation for the East Asian-U.S. parental involvement gap presented in this chapter shifts focus from national cultural beliefs to schools. Case studies of preschools and elementary schools in Japan contend that the Japanese educational system, through formal school practices to engage parents in the schooling process, prompt parents who might otherwise take a backseat to their children's education to become involved. Sociologists provide a theoretical basis to understand why schools might make a difference in the proportion of involved parents. Tests of the second, third, and fourth hypothesis—that the likelihood of involvement varies across schools within each nation after controlling for family background; that the quality of family-school relationships influences the likelihood of involvement, and that concentrated disadvantage in this school community decreases the likelihood of involvement— give us some indication of the relevance of this argument. To empirically address these hypotheses, I predict whether or not parents provide a desk at home using two-level, random-intercept logistic regression models. These models allow examination of relationships between features of the school community and provision of a desk at home while holding constant parent background factors and student characteristics. This method also allows me to assess the extent to which the likelihood that parents provide a desk at home varies across schools.

In these models, the school-specific random intercept ζ_j , which has a mean of zero and variance ψ, is in the linear predictor of provision of a desk at home. I assume that underlying the observed dichotomous response y_{ij} (whether or not family *i* in school *j* had provided their fourth grader with a study desk at home) is an unobserved or latent continuous response y_{ij}^*, representing the propensity of family *i* in school *j* to provide a desk at home (Rabe-Hesketh and Skrondal 2008). The residual error term associated with each family, ϵ_i, is assumed to have a standard logistic distribution with a mean of zero and variance $\pi^2/3 \approx 3.29$. By conceptualizing the binomial outcome as a latent-response variable, I am able to quantify the between-school heterogeneity as the residual intraclass correlation, rho, or the proportion of the total variation in the propensity of families to provide a desk that is attributed to schools (Rabe-Hesketh and Skrondal 2008; Raudenbush and Bryk 2002). The equation for calculating the residual intraclass correlation is:

$$\text{rho} = \psi / (\psi + \pi^2/3)$$

To address the second hypothesis, that the likelihood of involvement varies across schools within each nation even after controlling for parents' background, I model provision of a desk at home using measures of parents' background and student characteristics. I call this "Model 1." The rho value of this model addresses the second hypothesis by telling us the proportion of variation in the propensity to provide a desk that is explained by differences between schools within each nation. The statistical software program Stata provides sensitivity tests of rho by testing the null hypothesis that $\psi = 0$ (that there is no between-school heterogeneity in parental involvement) with likelihood-ratio tests, and provides a p-value based on the correct sampling distribution of the test statistic (Rabe-Hesketh and Skrondal 2008). Tests of the null hypothesis that result in small p-values can be rejected at standard significance levels, and confirm that the likelihood that parents provide a desk at home varies significantly across schools.

To test the third and fourth hypothesis, that family-school relationships and concentrated disadvantage are positively and negatively associated, respectively, with the likelihood of parental involvement, I create another model, Model 2, which adds to Model 1 indicators of the quality of family-school relationships according to school staffs' reports, the level of concentrated disadvantage in the school community, the degree of neighborhood violence in the school community according to teachers, and other school-level controls.

Results from these analyses are presented in Table 4.3. The table is separated into five vertical panels, one for each nation. The first three panels provide results for the U.S., Hong Kong and Japan. The last two panels provide results for Singapore and Taiwan. For each nation, the results of Model 1 are displayed in the first column of data and results for Model 2 are in the second column. The coefficients are given as estimated log odds ratios, which can be interpreted as the difference in log-odds associated with a unit change in the corresponding variable. For ease of interpretation, I standardize all variables that are not dichotomous to have means equal to the national mean and a standard deviations of one. In this way, the coefficients explain the effect associated with a one standard deviation increase from the national mean. Following Lee and Burkam (2003), I denote relationships that are statistically significant at the $p < 0.10$ level in addition to the 0.001, 0.01 and 0.05 levels, because the reliability of multilevel relationships is influenced by the within-school sample size, which is considerably smaller than the full sample.[38] I limit my description of the results of Model 1 to the rho values and to the coefficients for family-school relationships and concentrated disadvantage in Model 2, as these are the focus of my analyses.

According to the results presented in Table 4.3, the rho values of Model 1, which are located in the third-to-last row of the first column of data for each nation, indicate that the proportion of the total variance associated with differences between schools is small (0.045 in the U.S. sample, 0.035 in the Hong Kong sample, 0.047 in the Japan sample, 0.044 in the Singapore sample and 0.023 in the Taiwan sample) but statistically significant (at the $p < 0.001$ level) in every nation. This finding suggests that the likelihood that parents from similar backgrounds provide a desk at home depends to some extent on their children's school. This evidence is consistent with the second hypothesis; it tells us that the likelihood of providing a desk at home varies statistically significantly across schools within each nation.

[38] The average number of families clustered in each school within each nation is sufficient for multilevel estimations (32 in the U.S., 31 in Hong Kong, 28 in Japan, 32 in Singapore and 29 in Taiwan).

Table 4.3 Two-Level Logistic Regression of Provision of a Desk at Home

	United States		**Hong Kong**		**Japan**	
Variables	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
2003 TIMSS	-0.104^	-0.085	-0.084	-0.071	0.047	-0.046
	(0.061)	(0.052)	(0.077)	(0.076)	(0.113)	(0.117)
Family Background						
Low SES	-0.776***	-0.688***	-0.513***	-0.474***	-0.563***	-0.552***
	(0.059)	(0.059)	(0.071)	(0.071)	(0.130)	(0.130)
High SES	0.537***	0.485***	0.466***	0.421***	0.265^	0.257^
	(0.075)	(0.075)	(0.096)	(0.096)	(0.141)	(0.141)
Non-native Language	-0.111^	-0.051	0.096	0.106	-1.000***	-0.996***
	(0.067)	(0.067)	(0.072)	(0.072)	(0.282)	(0.281)
Immigrant Status	-0.166**	-0.118*	-0.280***	-0.230***	-0.747***	-0.715**
	(0.052)	(0.053)	(0.062)	(0.063)	(0.221)	(0.221)
Student Characteristics						
Age	-0.053*	-0.043^	-0.108***	-0.088**	-0.122*	-0.122*
	(0.023)	(0.022)	(0.029)	(0.029)	(0.048)	(0.048)
Girl	0.407***	0.417***	0.067	0.063	0.433***	0.430***
	(0.045)	(0.045)	(0.059)	(0.059)	(0.102)	(0.102)
School Community						
Staff Reports of Parental Support		0.110**		0.080^		0.157**
		(0.043)		(0.046)		(0.061)
Concentrated Disadvantage		-0.169***		-0.110*		0.026
		(0.039)		(0.045)		(0.060)
Neighborhood Violence		-0.125***		-0.042		-0.026
		(0.038)		(0.039)		(0.058)
% Non-native Lang.		0.012		0.126*		-0.010
		(0.032)		(0.055)		(0.065)
School Size		0.007		0.073^		0.068
		(0.030)		(0.041)		(0.061)
Large Community		0.102		0.133		-0.348*
		(0.069)		(0.086)		(0.135)
Small Community		-0.090		0.026		0.055
		(0.064)		(0.270)		(0.262)
Constant	1.267***	1.215***	1.194***	1.064***	2.783***	3.056***
	(0.053)	(0.060)	(0.075)	(0.099)	(0.101)	(0.145)
rho	0.045	0.020	0.035	0.020	0.047	0.033
N of Students	12,640	12,640	6,339	6,339	7,747	7,747
N of Schools	393	393	206	206	281	281

*** p<0.001, ** p<0.01, * p<0.05, ^ p<0.10; Standard errors in parentheses.

Table 4.3 Continued

	Singapore[a]		Taiwan	
Variables	Model 1	Model 2	Model 1	Model 2
2003 TIMSS	0.227**	0.226**	0.156*	0.170*
	(0.078)	(0.074)	(0.077)	(0.073)
Family Background				
Low SES	-0.823***	-0.786***	-0.838***	-0.799***
	(0.083)	(0.082)	(0.081)	(0.081)
High SES	0.308**	0.276**	0.530***	0.505***
	(0.101)	(0.101)	(0.119)	(0.119)
Non-native Language	-0.664***	-0.605***	-0.281***	-0.239**
	(0.072)	(0.072)	(0.081)	(0.081)
Immigrant Status	0.063	0.122^	-0.195*	-0.177*
	(0.067)	(0.068)	(0.087)	(0.087)
Student Characteristics				
Age	0.028	0.038	-0.048	-0.048
	(0.031)	(0.031)	(0.034)	(0.034)
Girl	0.212**	0.221***	0.058	0.060
	(0.066)	(0.065)	(0.068)	(0.068)
School Community				
Staff Reports of Parental Support		0.126**		0.051
		(0.047)		(0.040)
Concentrated Disadvantage		-0.042		-0.093*
		(0.045)		(0.043)
Neighborhood Violence		0.019		-0.099**
		(0.038)		(0.038)
% Non-native Lang.		-0.114**		-0.055
		(0.039)		(0.039)
School Size		0.083*		0.075^
		(0.039)		(0.045)
Large Community		-		-0.031
				(0.089)
Small Community		-		0.170
				(0.248)
Constant	2.358***	2.305***	2.048***	2.030***
	(0.084)	(0.083)	(0.073)	(0.094)
rho	0.044	0.028	0.023	0.010
N of Students	10,447	10,447	8,193	8,193
N of Schools	331	331	286	286

*** p<0.001, ** p<0.01, * p<0.05, ^ p<0.10; Standard errors in parentheses.

[a] The Singaporean sample does not include indicators of community size because principals' responses were "gang punched" to "more than 500,000 people."

Evidence consistent with Hypothesis 3, that high quality, family-school relationships act to increase the proportion of parents who are involved in their children's education, would be a positive association between the quality of family-school relationships, measured as staff characterizations of parental involvement and support, and the likelihood of provision of a desk at home. The results of Model 2, which are displayed in second column of data for each nation in Table 4.3, indicate that the relationship is in the expected direction in all nations. Staff characterizations of parental involvement and support for children's education are associated with an increase in the estimated log odds of parental provision of a desk at home in all nations, and statistically significantly so in all nations except Taiwan. An increase of one standard deviation in staffs' characterizations of parental involvement and support at their schools is associated with a percentage increase in the estimated odds of provision of a desk that is largest in Japan, at 17 percent (=100x(1-exp(0.157)), followed by Singapore (13%), the U.S. (12%), Hong Kong (8%) and Taiwan (5%). In other words, non-immigrant Japanese parents of average SES background whose son attends an average elementary school in a medium-size community with high-quality family-school relationships are 17 percent more likely than their counterparts at a similar school with average family-school relationships to provide their son with a desk at home.

According to Hypothesis 4, the Disadvantage scale variable should be negatively related to provision of a desk at home. The results, which are displayed in the second column of data for each nation in Table 4.3 demonstrate that, consistent with expectations, Disadvantage is negatively associated with provision of a desk in the U.S., Hong Kong, Singapore and Taiwan, although the relationship is not statistically significant in Singapore. The size of the estimated log odds ratio for Disadvantage is greatest in the U.S. (-0.17, statistically significant at the p<

0.001 level) followed by Hong Kong (-0.11, p<0.05 level), Taiwan (-0.09, p<0.05 level) and Singapore (-0.04, not statistically significant). For example, native-born, English-speaking parents of average SES background whose son attends an elementary school in a medium-sized, disadvantaged community are 16 percent less likely than their counterparts whose son attends an elementary school with an average score on the Disadvantaged scale to provide a desk at home. In Japan, the relationship between disadvantage in the school community and provision of a desk at home is positive, but the estimated standard error for the effect is more than twice the size of the coefficient. The data do not provide a clear picture of the relationship between provision of a desk and the level of disadvantage in school communities in Japan.

Given that levels of violence tend to be greater in areas of concentrated disadvantage (Sampson et al. 1997), we might also expect to find a negative relationship between that measure and provision of a desk at home. The results in Table 4.3 are generally consistent with this claim. The Neighborhood Violence scale is negatively associated with provision of a desk at home in all nations except Singapore. The coefficient for Neighborhood Violence is largest in the U.S. (-0.13, statistically significant at the p<0.001 level) followed by Taiwan (-0.10, p<0.01 level) and Hong Kong (-0.04, not statistically significant). The coefficient for Neighborhood Violence is also negative in Japan, but the standard error is large relative to the coefficient, which suggests that the data do not provide a clear picture of the relationship. The same is true for the Singaporean data; the coefficient for Neighborhood Violence is positive but small relative to the size of the standard error, indicating that the data do not capture well the relationship between safety and provision of a desk at home.

Lastly, the inclusion of school characteristics in Model 2 explains some but not all of the between-school variance in provision of a desk at home in the U.S., Hong Kong, Japan and

Singapore. In these nations, the estimated rho value for Model 2 are smaller than in Model 1, but still statistically significant. From Model 1 to Model 2, the proportion of the total variance that is explained by differences between schools decreases by 56 percent ((0.045-0.020)/0.045)x100) in the U.S., 43 percent in Hong Kong, 30 percent in Japan, 36 percent in Singapore and 57 percent in Taiwan. For the Taiwan data, the lack of statistical significance of the rho value for Model 2 indicates that likelihood ratio tests of the null hypothesis that $\psi = 0$ cannot be rejected, or that the school-level variables added to Model 2 mostly explain the differences between schools in their relationships to parental provision of a desk at home. These findings provide some empirical support for the alternative argument that features of the school community, specifically the quality of family-school relationships and the degree of concentrated disadvantage, tend to influence parents' decisions to become involved in their children's education.

Why is the proportion of involved parents higher in parts of East Asia than in the U.S.?

Given both the demonstrated, negative relationship between concentrated disadvantage and mutual trust and respect (Sampson et al. 1997) and the importance of mutual trust and respect to parental involvement, it is possible that the proportion of parents who are involved in their children's education is lower in the U.S. and Hong Kong because there are more school communities characterized by concentrated disadvantage in those nations than in Japan, Singapore and Taiwan. To assess this hypothesis (Hypothesis 5), I use the Disadvantage scale to classify schools into three categories: Affluent, Average and Disadvantaged. Schools in the Affluent category have scores on the Disadvantage scale that are one standard deviation or more below the mean. Schools with scores on the Disadvantage scale that are within one standard deviation of the mean are categorized as Average, while schools with scores that are one standard deviation or more above the mean are categorized as Disadvantaged.

To compare the extent to which levels of disadvantage in school communities differ across nations, I calculate the mean scores using population weights for each category of schools on a number of features of the school community. For each feature of the school community, I test whether outcomes for Affluent and Disadvantaged schools differ from Average schools using two-tailed Wald tests. If hypothesis 5 is correct, then the results should indicate that in Hong Kong and the U.S., Disadvantaged schools typically have fewer resources and weaker relationships between parents and school staff than Average schools. Further, there should be less inequality in the socioeconomic conditions of Disadvantaged schools compared to Average and Affluent schools in Japan, Singapore and Taiwan compared to Hong Kong and the U.S.

Table 4.4 displays the results of these analyses. The table is divided into five panels, one for each nation. Each panel consists of three columns of data. The first column displays results for Affluent schools, the second column shows results for Average schools and the third column reports results for Disadvantaged schools. The first row of data in each panel provides the mean score on the Disadvantage scale for each category of schools. These results reveal cross-national variation in the difference between the proportion of economically disadvantaged and affluent students who attend Disadvantaged schools. Consistent with expectations, scores are higher in the U.S. (58.0) and Hong Kong (54.5) than in Japan (12.3), Singapore (37.8) and Taiwan (28.8). In fact, in Japan the mean for Disadvantaged schools on the Disadvantage scale is below the mean for Average schools in the U.S. (14.3) and roughly equal to the mean of Average schools in Hong Kong (11.7).

Table 4.4 Features of the School Community by Level of Concentrated Disadvantage

	United States			Hong Kong		
Variables	Affluent[a]	Average	Disad-vantaged	Affluent	Average	Disad-vantaged
Concentrated Disadvantage	-41.70***	14.29	58.00***	-47.71***	11.71	54.47***
% Disadvantaged	15.12***	29.76	63.00***	9.04***	25.48	63.00***
% Advantaged	56.82***	15.47	5.00***	56.75***	13.77	8.53***
Neighborhood Violence	1.20***	1.45	1.98***	1.57	1.68	1.91**
School Resources	1.63***	1.97	1.90**	1.61*	1.79	2.12***
Parent-School Relationship	4.17***	3.40	2.59***	3.74***	3.25	2.97***
% High quality Relations	77.39***	33.92	2.57***	70.47***	22.22	13.60***
% Low quality Relations	0.00**	6.33	33.74***	0.00*	4.64	13.62*
% Non-native Language	9.42	12.38	25.58***	7.26*	5.10	5.00*

	Japan			Singapore		
Variables	Affluent	Average	Disad-vantaged	Affluent	Average	Disad-vantaged
Concentrated Disadvantage	-58.00***	-31.91	12.25***	-55.65***	-8.45	37.76***
% Disadvantaged	5.00***	8.63	21.65***	7.35***	10.31	45.80***
% Advantaged	63.00***	40.54	9.40***	63.00***	18.76	8.04***
Neighborhood Violence	1.95	1.97	1.99	1.50	1.56	1.65*
School Resources	1.73	1.76	1.83	1.35	1.36	1.32
Parent-School Relationship	3.33	3.32	3.00**	3.68***	3.27	2.62***
% High quality Relations	28.49	24.98	9.18*	46.36***	19.22	0.00***
% Low quality Relations	2.44	2.95	11.07	0.00**	2.30	28.36***
% Non-native Language	5.00	5.53	5.24	2.46***	3.41	3.65***

	Taiwan		
Variables	Affluent	Average	Disad-vantaged
Concentrated Disadvantage	-44.26***	-3.21	28.84***
% Disadvantaged	5.90***	8.02	37.01***
% Advantaged	50.16***	11.23	8.17***
Neighborhood Violence	1.92	1.86	1.82
School Resources	2.10	2.10	2.28
Parent-School Relationship	4.14***	3.86	3.48***
% High quality Relations	88.78**	70.77	50.70***
% Low quality Relations	0.00	0.00	2.53
% Non-native Language	16.85***	26.41	30.77**

*** p<0.001, ** p<0.01, * p<0.05, ^ p<0.10; Significance levels indicate that outcomes differ statistically significantly from outcomes for Average schools.

The second and third rows of data in each panel help to explain why mean scores on the Disadvantaged scale for Disadvantaged schools differ across nations. On average, 63 percent of the students enrolled in Disadvantaged schools in the U.S. and Hong Kong are from economically disadvantaged backgrounds, while only 5.0 percent and 9.0 percent of students in Disadvantaged schools in the U.S. and Hong Kong, respectively, come from economically affluent families. In Japan, Singapore and Taiwan, at most 7.4 percent of the student population of Disadvantaged schools come from economically affluent families. The difference between Hong Kong and the U.S., on the one hand, and Japan, Singapore and Taiwan on the other, is in the percentage of students who come from economically disadvantaged families and attend Disadvantaged schools. While over 50 percent of students attending Disadvantaged schools in the U.S. and Hong Kong are from disadvantaged families, only 21.7 percent, 37.0 percent and 45.8 percent of students attending Disadvantaged schools in Japan, Taiwan and Singapore, respectively, are from economically disadvantaged families.

The next row of data in each panel of Table 4.4 reports the mean for each school category on the Neighborhood Violence scale. Here we see that within the U.S. and Hong Kong, outcomes for Disadvantaged schools differ statistically significantly from outcomes for Average schools. Disadvantaged schools in the U.S. average a score of 2.0 on the Neighborhood Violence scale, which is statistically significantly greater than the mean of 1.5 for Average schools and 1.2 for Affluent schools. In Hong Kong, the mean score of 1.9 on the Neighborhood Violence scale for Disadvantaged schools is statistically significantly higher than the mean of 1.7 for Average schools. The mean score on the Neighborhood Violence scale for Disadvantaged schools in Singapore is also statistically significantly higher than the mean for Average schools (1.65 versus 1.56). On the other hand, in Japan and Taiwan, differences across categories of schools in mean scores on the Neighborhood Violence scale are not statistically significant. These findings suggest that Disadvantaged schools tend to be located in less safe neighborhoods in the U.S. and Hong Kong and Singapore, but not in Japan and Taiwan.

To assess differences across categories of schools within each nation in the quantity of resources available to aid instruction, I created a scale variable based on a series of survey items directed to school principals on the extent to which a shortage of 18 different resources affects their school's capacity to provide instruction.[39] Responses could range from 1 to 4 indicating responses of "None," "A little," "Some," or "A lot." Higher average scores on this scale indicate greater shortages of resources that aid instruction. Mean scores on this Resource Shortage scale for each category of schools are reported on the fifth row of data in each panel of Table 4.4. In Japan, Singapore and Taiwan, mean scores on the Resource Shortage scale do not differ statistically significantly across Affluent, Average and Disadvantaged schools. The same cannot be said for the U.S. and Hong Kong. In Hong Kong, the mean for Resource Shortage is statistically significantly higher for Disadvantaged schools than for Average schools. In the U.S., the mean score for Disadvantaged schools on the Resource Shortage scale is not statistically significantly different from the mean for Average schools. However, the mean for Affluent schools is statistically significantly smaller than the mean for Average schools, indicating that, compared to Affluent schools, both Average and Disadvantaged schools in the U.S. tend to lack resources that aid instruction.

[39] The list of resources is: 1) Instructional materials (e.g. textbook), 2) Budget for supplies (e.g. paper, pencils), 3) School buildings and grounds, 4) Heating/cooling and lighting systems, 5) Instructional spaces (e.g. classrooms), 6) Special equipment for handicapped students, 7) Computers for math instruction, 8) Computer software for math instruction, 9) Calculators for math instruction, 10) Library materials relevant to math instruction, 11) Audio-visual resources for math instruction, 12) Science laboratory equipment and materials, 13) Computers for science instruction, 14) Computer software for science instruction, 15) Calculators for science instruction, 16) Library materials relevant to science instruction, 17) Audio-visual resources for science instruction, and 18) Teachers.

Hypothesis 5 rests on the assumption that concentrated disadvantage in the school community negatively impacts the likelihood that parents become involved in their children's education because it stymies efforts to develop parent-school relationships based on mutual trust and respect. Evidence consistent with this assumption would show that school staffs' characterization of parental involvement and support are lowest in Disadvantaged schools within each nation. The last three rows of data in each panel of Table 4.4 allow us to examine this assumption. Consistent with expectations, outcomes on the variable measuring staff reports of parental support are statistically significantly lower in Disadvantaged schools than in Average schools in every nation. Additionally, in every nation except Japan, mean scores on the Parent-School Relations scale are statistically significantly higher in Affluent schools than in Average schools, suggesting that in these nations, school staffs' characterizations of parental support and involvement increase with the socioeconomic status of the school community. The fact that staff characterizations of parental involvement and support do not differ greatly by school SES in Japan suggests, perhaps, that the nation's school-based outreach programs are also positively influencing staff perceptions of parental involvement, such that they are grounded less on parents' socioeconomic status and more on what parents are doing to help their children succeed in school.

To gain a more complete picture of the range in the quality of parent-school relationships at Affluent, Average and Disadvantaged schools within each nation, the second-to-last row of data in each panel of Table 4.4 shows the percentage of schools in each category (e.g. Affluent, Average and Disadvantage) whose staff characterized parental support and involvement at their school as "high" or "very high," while the last row of data in each panel indicate the percentage of schools in each category whose staff characterize parental support and involvement at their schools as “low” or “very low.” In each nation, there is a statistically significant difference across categories of schools in the percentage with high quality parent-school relationships (as measured by staffs’ characterization of parental support and involvement at their schools). In all nations, the percentage of schools with high quality parent-school relationships is lowest in the Disadvantaged category. For example, in the U.S. 77.4 percent of Affluent schools characterized parental support and involvement as high compared to 33.9 percent of Average schools and only 2.6 percent of Disadvantaged schools. In Singapore, not a single Disadvantaged school characterized parental support and involvement as high.

Interestingly, in every nation except Japan staff at Affluent schools did not characterize parental support or involvement as low. Zero Affluent schools in the U.S., Hong Kong, Singapore and Taiwan have low-quality family-school relationships (as measured by staff characterization of parental support and involvement)—and the outcomes are statistically significantly different from the outcomes for Average schools within these nations. In Japan, 2.4 percent, or one, Affluent school received a characterization of low parental support and involvement, and that percentage does not differ statistically significantly from those of Average nor Disadvantaged schools. This finding sets Japan apart from the rest of the sample and suggests that school staff in Japan rarely characterize parental support and involvement at their schools as low. In contrast, parent-school relationships are characterized as low in 33.7 percent of Disadvantaged schools in the U.S.

In sum, the results presented in Table 4.4 indicate greater inequalities in the socioeconomic conditions of school communities in Hong Kong and the U.S. in contrast to Japan, Singapore and Taiwan. Disadvantaged schools in Hong Kong and the U.S. enroll a greater percentage of economically disadvantaged students compared to Disadvantaged schools in

Japan, Singapore and Taiwan; they are located in less safe neighborhoods than Affluent or Average schools, and have fewer resources that aid instruction than Average or Affluent schools. Additionally, consistent with the argument that concentrated disadvantage in the school community negatively impacts parents' decisions to become involved in their children's education because it stymies the development of social relationships based on mutual trust and respect, the results reported in Table 4.4 show a negative relationship between parent-school relationships (as measured by staff characterizations of parental support and involvement) and the level of disadvantage in school community in every nation under investigation. In the next section I explore the extent to which staff characterizations of parental involvement and support (or the quality of parent-school relationships) depend on the proportion of parents who are involved in their children's education at their school (as measured by provision of a desk at home) versus the degree of disadvantage in the school community.

Predicting staff characterizations of parental support

According to Hypothesis 6, the proportion of parents involved in their children's education in the U.S. is lower than in Japan and Taiwan because educators in the U.S. perceive of parental involvement as determined by parents' social-class background as opposed to their own actions and efforts to engage parents in the schooling process. Evidence consistent with Hypothesis 6 would show that staff characterizations tend to be less positive of parental support and involvement, or the quality of family-school relationships tends to be lower, in Disadvantaged schools compared to Average schools holding other factors constant. Additionally, we should find a positive relationship between Affluent schools and staff characterizations of parental involvement and support, and a negative relationship between the level of violent and criminal activity in the school neighborhood (as measured by teachers' reports of feeling safe at school and in the surrounding locale). Further, indicators of the degree of disadvantage in the school community should explain more of the variation in staff characterizations of parental involvement and support, or family-school relationships, in the U.S. than in other East Asian nations.

To empirically assess this hypothesis, I use OLS regression to model staff characterizations of parental involvement and support, or the quality of parent-school relationships, using school-level data from each nation. The first model examines the relationship between the proportion of parents in the school who had provided their children with a desk at home and staff characterizations of parental involvement and support controlling for school size, the size of the school community and the TIMSS round. In Model 2 I include the Neighborhood Violence scale variable and dummy variables indicating that schools are either Affluent or Disadvantaged. I rely on the dummy variables as opposed to the Disadvantaged scale variable to allow the effect of disadvantage on staff characterizations of parental support and involvement to be nonlinear. Again, to ease interpretations of the relationships between the dependent and independent variables, I standardized the independent variables that are not dichotomous so that the national mean is equal to zero and it has a standard deviation of one.

Table 4.5 OLS Regression of Staff Characterizations of Parental Support

	United States		**Hong Kong**		**Japan**		**Singapore**		**Taiwan**	
Variables	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
2003 TIMSS	0.019	0.042	0.141	0.199**	0.174**	0.280***	-0.069	-0.066	-0.085	-0.087
	(0.072)	(0.059)	(0.073)	(0.071)	(0.065)	(0.068)	(0.056)	(0.052)	(0.056)	(0.053)
School Size	0.067	0.003	0.127**	0.085*	0.034	0.043	0.126***	0.078**	0.080*	0.076**
	(0.036)	(0.033)	(0.041)	(0.038)	(0.034)	(0.034)	(0.031)	(0.030)	(0.032)	(0.029)
Large Community	-0.074	0.119	-0.055	-0.036	-0.050	-0.063			0.091	0.067
	(0.096)	(0.079)	(0.088)	(0.084)	(0.077)	(0.075)			(0.068)	(0.066)
Small Community	-0.096	-0.140*	0.084	0.003	-0.040	-0.123			-0.255	-0.194
	(0.086)	(0.071)	(0.144)	(0.092)	(0.112)	(0.108)			(0.166)	(0.163)
% Non-native Lang.	-0.162***	-0.026	0.006	-0.036	0.038*	0.034**	-0.097**	-0.067*	-0.022	0.004
	(0.045)	(0.037)	(0.038)	(0.038)	(0.017)	(0.013)	(0.031)	(0.028)	(0.027)	(0.027)
% w/ Desk at Home	0.461***	0.170***	0.186***	0.126***	0.114***	0.097**	0.165***	0.126***	0.102***	0.071*
	(0.043)	(0.043)	(0.038)	(0.036)	(0.033)	(0.031)	(0.027)	(0.025)	(0.030)	(0.028)
Affluent School		0.569***		0.340***		0.143		0.306***		0.228***
		(0.075)		(0.083)		(0.104)		(0.065)		(0.066)
Disadvantaged School		-0.359***		-0.079		-0.304***		-0.423***		-0.275*
		(0.077)		(0.100)		(0.070)		(0.091)		(0.108)
Neighborhood Violence		-0.302***		-0.132***		-0.081**		-0.062		-0.076*
		(0.045)		(0.039)		(0.031)		(0.032)		(0.030)
Constant	3.281***	3.240***	3.287***	3.196***	3.225***	3.225***	3.285***	3.265***	3.868***	3.860***
	(0.070)	(0.068)	(0.091)	(0.092)	(0.071)	(0.070)	(0.036)	(0.038)	(0.063)	(0.062)
Adjusted R-squared	0.327	0.559	0.182	0.298	0.065	0.143	0.226	0.341	0.124	0.196
Change in R-squared	0.232		0.116		0.078		0.115		0.072	
Observations	393	393	206	206	281	281	331	331	286	286

*** p<0.001, ** p<0.01, * p<0.05; Robust standard errors are in parentheses.

Table 4.5 presents the results from these analyses. Table 4.5 is divided into five panels, with the results for each nation displayed in a separate panel. For each nation, results of Model 1 are located in the first column of data and results for Model 2 are in the second column. Huber-White standard errors are given in parentheses below the coefficients for each variable. As expected, the results of Model 1 indicate a statistically significant, positive relationship at the school level between the proportion of parents at school who have provided their fourth grader with a desk at home and staff characterizations of parental support and involvement (or the quality of parent-school relationships). An increase of one standard deviation above the mean in the proportion of parents who provided their fourth grader with a desk at school is associated with an increase in staff characterizations of parental support and involvement of 0.50 in the U.S., 0.19 in Hong Kong, 0.12 in Japan, 0.19 in Singapore and 0.10 in Taiwan.

The adjusted R-squared for Model 1 indicates that the proportion of parents who provide a desk at home and the controls together explain at most 30 percent of the variation in staff characterizations of parental involvement (in the U.S.) and as little as 7 percent (in Japan). Together these findings suggest that staff characterizations of parental involvement and support, or the quality of parent-school relationships, is positively related at the school level to the proportion of involved parents.

The inclusion of indicators of concentrated disadvantage in the school community in Model 2 shed light on the relevance of Hypothesis 6 for addressing the East Asian-U.S. parental involvement gap. Consistent with Hypothesis 6, the results of Model 2, which are reported in the second column of data of each panel of Table 4.5, demonstrate a negative relationship between staff characterizations of parental involvement and Disadvantaged schools, net of other factors. The estimated negative effect of concentrated disadvantage on staff characterizations of parental involvement is largest, on average, in Singapore (-0.45, statistically significant at the $p<0.001$ level) followed by the U.S. (-0.36, $p<0.001$ level), Japan (-0.32, $p<0.001$ level), Taiwan (-0.28, $p<0.05$ level) and Hong Kong (-0.08, not statistically significant). Staff characterizations of parental involvement, or the quality of family-school relationships, also tend to be more positive at Affluent schools than at Average schools. The estimated effects of working at an Affluent school on staff characterizations of parental support and involvement, with all else held constant, is on average largest in the U.S. (0.58, statistically significant at the $p< 0.001$ level) followed by Hong Kong and Singapore (0.32 and statistically significant at the $p< 0.001$ level in both nations), Taiwan (0.23, $p<0.001$ level) and Japan (0.15, not statistically significant). This finding is consistent with Lareau's (2000[1989]): school personnel in the U.S. tend to equate levels of parental involvement to parents' beliefs about the importance of education and social-class background, as opposed to parents' lack of knowledge or to a failure to communicate.

The last indicator of disadvantage in the school community, the degree of violent and criminal activity in the neighborhood (as measured by teachers' reports of feeling safe at school and in the surrounding neighborhood), is also negatively related to staff characterizations of parental support and involvement, or the quality of family-school relationships. A one standard-deviation increase on the Neighborhood Violence scale is associated with a decrease in staff characterizations of parental support and involvement, holding all else constant. Moreover, the estimated effect, which is on average largest in the U.S. (-0.31, statistically significant at the $p<0.001$ level), is statistically significant in all nations except Singapore. Taken together these findings provide some support for the argument that staff characterizations of parental involvement, or the quality of family-school relationships, are based to some extent on the degree of disadvantage in the school community in *all* nations.

However, for Hypothesis 6 to be relevant to understanding the East Asian-U.S. parental involvement gap, the socioeconomic status of the school community must explain more of the variation in staff characterizations of parental involvement and support (or the quality of family-school relationships) in the U.S. than in Japan, Singapore or Taiwan. The last row of data displayed in each nation's panel in Table 4.5 report the change in the adjusted R-squared that is due to the inclusion of indicators of disadvantage in Model 1. These figures indicate the proportion of the total variance in staff characterizations of parental involvement and support, or the quality of family-school relationships, that can be attributed to the degree of disadvantage in school communities. Consistent with Hypothesis 6, concentrated disadvantage in school communities explains a much larger portion of the variance in staff characterizations of parental involvement and support in the U.S. (0.256) than in any East Asian nation. (In Singapore, the amount of variance explained by socioeconomic conditions is 0.139 followed by Hong Kong, 0.112, Japan, 0.074, and Taiwan, 0.073.) To put this finding into perspective, the amount of variation in staff characterizations of parental involvement that is explained by disadvantage in school communities in the U.S. sample is nearly twice as great as in the Singapore sample, and three and one-third times the amount in the Taiwan sample.

Additional evidence that staff characterizations of parental involvement and support, or the quality of family-school relationships, are more strongly influenced by the degree of disadvantage in the school community in the U.S. than in East Asia comes from post-estimation significance tests on the difference in the coefficient for parental provision of a desk in Model 1 and Model 2. As Table 4.5 reports, the estimated coefficient for the percentage of students with a desk at home at each school is smaller in every nation in Model 2 than in Model 1. If the difference is statistically significant, then this finding suggests that the relationship between the proportion of parents in the school community who are involved in their children's education (as measured by parental provision of a desk at home) and staff characterizations of parental involvement and support (or the quality of family-school relationships) is spurious. Disadvantage in the school community explains some of the positive association between parental provision of a desk at home and staff characterizations of parental involvement or support (or the quality of family-school relationships). Wald tests that the coefficients for the percentage of students with a desk at home are the same in Model 1 and Model 2 indicate that the null hypothesis cannot be rejected in any of the East Asian nations, but that it can be rejected in the U.S. (at the $p < 0.001$ level). It appears as though staff characterizations of parental involvement in the U.S. are more influenced by the degree of disadvantage in the school community than in East Asian nations.

CONCLUSION

Using data from all high-achieving nations that participated in the 2003 or 2007 TIMSS fourth-grade population, this chapter shows that while the proportion of parents who provide a desk at home is still higher in Japan and Taiwan than in the U.S. twenty years after Stevenson and Stigler (1992) publicized that finding, there is considerable variation within East Asia in the proportion of parents who are involved in their children's education in this manner. The proportion of parents who provide their children with a desk at home is very high (although not uniquely so) in Japan, around the international mean in Taiwan and Singapore, and lower in Hong Kong than in the U.S.

This chapter presents empirical evidence consistent with the argument family-school relationships and the degree of concentrated disadvantage in the school community influence parents' decisions to become involved in their children's education. The findings show that, after controlling for parent background factors and student characteristics, the likelihood that parents provide a desk at home varies across schools within each East Asian nation and the U.S. In all nations, staff characterizations of parental involvement and support, which were used to indicate the quality of family-school relationships in the school community, are positively related to provision of a desk at home. In all nations but Taiwan, the degree of concentrated disadvantage in the school community is negatively related to parental provision of a desk at home. Including these variables into the model of parental provision of a desk at home explains most, if not all (in the case of Taiwan) of the variation between schools in the likelihood that parents provide their children with a desk at home.

This chapter develops two hypotheses based on the alternative theory to explain the East Asian-U.S. parental involvement gap. The first, based on social organizational research, is that the degree of concentrated disadvantage in the school community negatively impacts the formation of high quality parent-school relationships. Consistent with this hypothesis, I show that the degree of concentrated disadvantage is greater in Hong Kong and the U.S. than in Japan, Singapore and Taiwan. Schools with the least affluent student populations in Hong Kong and the U.S. enroll a greater percentage of economically disadvantaged students compared to similar schools in Japan, Singapore and Taiwan. Unlike similar schools in Japan, Singapore and Taiwan, where the differences between the least affluent schools and other schools in terms of resources and neighborhood safety are not significant, the least affluent schools in Hong Kong and the U.S. are located in less safe neighborhoods than other schools in those nations (according to teachers' reports of feeling safe at school and in the surrounding area), and they have fewer resources that aid instruction than other schools. Taken together, these findings suggest that there is more concentrated disadvantage in school communities in Hong Kong and the U.S. than in Singapore, Taiwan and Japan.

Further support for the hypothesis that concentrated disadvantage in the school community limits school staffs' ability to foster strong bonds with parents, the analyses of this chapter reveal a negative association between parent-school relationships (as measured by staff characterizations of parental support and involvement) and socioeconomic disadvantage in school community in every nation under investigation. Perhaps future research could shed more light on the relationship between concentrated disadvantage and parental involvement by expanding the scope of the analysis to include a larger number of nations with high levels of socioeconomic inequality across schools.

The second hypothesis developed in this chapter to explain the East Asian-U.S. parental involvement gap is that educators in the U.S. attribute parents' involvement to social-class background more so than educators in East Asian nations. Consistent with this argument, the analyses of this chapter demonstrate that in the U.S., staff characterizations of parental involvement and support are strongly influenced by the socioeconomic status of the school community. Whether or not schools enroll the most affluent or least affluent student populations and an indicator of neighborhood violence account for more than 25 percent of the variation in staff characterizations of parental involvement and support in the U.S., compared to 7.6 percent in Taiwan, 8.5 percent in Japan, 11.3 percent in Hong Kong and 12.8 percent in Singapore. Additionally, in contrast to any East Asian nation, in the U.S., some of the relationship between staff characterizations of parental involvement and the proportion of parents at school who provide their children with a desk is spurious.

Although this chapter underscores the importance of school staffs' perceptions for increasing parental involvement, the purpose of this chapter is not to exchange one cultural deprivation theory for another. Rather, I seek to highlight structural features in the U.S. educational system that currently reinforce staff perceptions of parental involvement. For example, I maintain that engaging parents in the schooling process is a skill that requires professional training. According to Epstein (1995), advanced degree programs for teachers in the U.S. do not typically emphasize parent-teacher partnerships in their curricula. State and local school districts could implement professional training and development programs for teachers to learn how to design and execute successful parental involvement programs, but this does not appear to be happening on a widespread basis. It is highly likely that staff attribute parental involvement to factors outside their control because it is difficult to increase parental involvement and because state and local school districts, as well as advanced degree programs for teachers, do not teach them how to do it.

To implement successful parental involvement programs, what parents do in their own homes with their own children needs to be considered to some degree by school staff as part of their professional responsibility. What U.S. scholars have failed to note is that the legal structure of the U.S. does not always encourage this kind of perspective, which could also be hindering school staffs' ability to increase parental involvement here. Over the last six decades, U.S. Supreme court rulings have limited schools' ability to act *in loco parentis* in part by augmenting students' rights to free speech. Scholars trace this trend back to *West Virginia State Board of Education v Barnette* (1943), which ruled that schools cannot force students to recite the pledge of allegiance. In the decision, the court reversed the common understanding of public schools as institutions geared at inculcating a sense of national unity and shared values and instead concluded that the U.S. public school system should cultivate rich cultural diversities. The decision clearly states that no government official, school teachers included, has the right to dictate the standards of appropriate behavior. While this legal structure does not preclude parent-teacher partnerships, it adds another layer to an already difficult task. Future cross-national research might ask how variations in the legal structure regarding schools' rights to act *in loco parentis* influence the mechanisms through which schools encourage parental involvement.

This chapter reveals the multiplex nature of the concept of parental involvement and emphasizes the importance of social resources adhering in the relationships between parents and school staff to increase the proportion of involved parents in school communities. However, appropriate measures of all of these important variables are lacking from internationally comparable datasets on early education, and TIMSS is no exception. More progress on understanding the mechanisms supporting higher proportions of involved parents in East Asian nations compared to the U.S. will be difficult until internationally-comparable datasets gather high quality data on these factors.

This chapter also demonstrates that to gain an accurate assessment of the number of parents who are involved in their children's education or the effectiveness of their involvement, and to have the ability to persuasively compare this measure across nations, there are several important points that internationally comparative researchers should keep in mind. First, does the type of parental involvement being measured depend on individual schools or teachers to provide parents with the opportunity to engage in such an activity? The amount of time that parents spend helping their children with homework turns out to be a poor measure of involvement because it relies too heavily on the amount of homework that teachers assign to children. Other measures of parental involvement, such as volunteering at the school, could also pose a problem, as this type of involvement depends on the frequency with which schools afford parents the opportunity to volunteer. Second, is the type of parental involvement under consideration heavily influenced by individual children's needs and abilities? Third, is the type of parental involvement popular and age-appropriate in all societies under consideration? Lastly, does the measure rely on rated responses to survey questions? If so, national differences in responses biases need to be taken into account before the measure can be used to compare levels of parental involvement across nations.

Chapter 5

Conclusion

THE QUESTION OF WHY East Asian nations continue to out-educate us in math is far from settled. Policymakers and academics tend to focus more on culture than on demonstrable differences in the math education and support for learning that East Asian systems provide compared to the U.S., and few have taken advantage of the high-quality, internationally comparable datasets available to empirically investigate East Asian education. This is what sociologists can bring to the table.

Analysis of data from the 2003 and 2007 TIMSS fourth-grade population has shown that variation in the quantity of math instruction fourth-grade students receive is smaller in all East Asian educational systems than in the U.S.; that the minimum percentage of math topics students might be taught is higher in East Asian educational systems than in the U.S., and that there tends to be more agreement regarding the curriculum among teachers in East Asian educational systems than in the U.S. These findings suggest that, compared to the U.S., the quantity of educational opportunities are more equitably distributed among fourth-grade students in East Asian educational systems. Further, it appears from these data that East Asian educational systems are doing a better job than the U.S. of ensuring that all fourth graders have access to basic math knowledge.

Additionally, the analysis has shown that educational systems like the U.S., which allow more variation in the quantity of math instruction provided across same-grade classrooms and also allow teachers to cover only a small fraction of the math topics that students are typically taught, tend to have lower average math achievement scores than those that ensure a higher baseline of instruction and less variation in the quantity of instruction that students receive. While not conclusive, this evidence is consistent with the argument that the institutional arrangements in the U.S. that allow for greater variability in the math instruction provided might be partially responsible for the poor performance of U.S. students on international assessments in math.

It is still unclear whether uniformity of instruction *on its own* would increase average achievement or whether curriculum preparation, teacher training in the curriculum or other factors regarding the curriculum that were not included in the analyses of this dissertation but might positively impact both the uniformity of instruction and average achievement, are to credit for the positive relationship between uniformity of instruction and achievement. The analysis of this dissertation cannot rule out the possibility that experimentation and variation in instructional content, coupled with a rigorously designed core curriculum or some such, could also promote achievement.

The analysis has also shown that East Asian teachers are considerably more effective at motivating student achievement in math than teachers in the U.S. Among students who have received the same quantity of instruction, average achievement is statistically significantly higher in East Asian nations than in the U.S. This work has shown that the most effective fourth-grade math teachers in the U.S. are about as effective as the average teacher in Japan, and less effective than the average in Hong Kong, Singapore and Taiwan. These findings are consistent with the argument that the relative inability of U.S. teachers to encourage math achievement is a prime determinant of the East Asian-U.S. achievement gap.

The analysis has also indicated that increasing the amount of material that students are taught does not necessarily increase achievement. At the system level there does not appear to be a coherent relationship between quantity of instruction and achievement. Although average achievement varies little within East Asia, there is considerable variation within East Asia in the quantity of math topics on the TIMSS assessment that fourth-graders are taught. On average across rounds, students in Singapore are taught about 85 percent of the test topics, while students in Japan are taught around 58 percent. Students in Hong Kong and Taiwan receive instruction on roughly 75 percent of the TIMSS topics. In comparison, students in the U.S. are taught about as many TIMSS math topics as are students in Singapore. Further, there does not appear to be a relationship at the student level between quantity of instruction and achievement. Holding student and school characteristics constant, including the percentage of test topics that students have been taught into three-level hierarchical models of student achievement does not predict achievement in any nation under investigation. While the results of this analysis are not conclusive, they suggest that the relationship between quantity of instruction and achievement is more complicated than the idea that the more students are taught, the more they will know.

Multilevel models of student achievement have also shed light on the extent to which teacher effectiveness varies within each nation. These results demonstrate that teachers in Hong Kong and Singapore vary significantly in their effectiveness—even after the quantity of instruction they provide is taken into account. In fact, variation in teacher effectiveness is greater in those two East Asian nations than it is in the U.S. Thus, while Hong Kong and Singapore compared to the U.S. appear to have attained more uniformity of math instruction across classrooms, they do not appear to have achieved significant uniformity in the *quality* of instruction they provide. By contrast, teacher effectiveness varies little in Japan and Taiwan, where there is also significant uniformity of math instruction across same-grade classrooms. These findings suggest that Japan and Taiwan are closest to having realized one of the core goals of schooling, which is to provide equal educational opportunities to all students.

However, results from multilevel models are far from conclusive because the precision of the measures of teacher effectiveness that they provide depends on students and teachers being randomly assigned to schools and to classrooms within schools. Practices such as ability grouping and assigning more experienced teachers to higher ability classrooms can influence the amount of variation in teacher effectiveness these types of models measure. Although this dissertation attempts to control for the nonrandom assignment of students to classrooms by including a dichotomous variable indicating whether or not schools group fourth-grade students my ability in math, it cannot account for the nonrandom assignment of teachers to schools or to classrooms within schools. Further, there is no way with the present data to take into consideration parents' choice of schools for their children. For this reason, results of multilevel models of teacher effectiveness should not be considered decisive.

It is widely believed that one reason for the East Asian-U.S. achievement gap is that a greater proportion of parents in East Asia than in the U.S. are involved in their children's education, and that East Asian parents are more effectively involved. The release of the best-selling *Battle Hymn of the Tiger Mother* (Chua 2011) fueled popular opinion that Chinese mothers are unforgiving in their quest to raise A+ students while Western parents by comparison are lackadaisical about stimulating their children's talents and abilities. Given the popularity of using cultural stereotypes to explain why East Asian parents are more involved in their children's education than U.S. parents, it is time for an alternative theory and for a rigorous analysis of recent, nationally representative data to inform this debate.

Unfortunately, all internationally comparable datasets on early education, including TIMSS, lack multiple high-quality measures of parental involvement from the nations under investigation. Fortunately, there is one indicator of parental involvement available in the TIMSS dataset—parental provision of a study desk at home—that is useful for several reasons. First, it is the same measure that Stevenson and Stigler (1992) draw on, which allows me to verify their results using nationally representative data. Additionally, provision of a desk at home is a good measure of parental involvement for internationally comparative research. This is because the measure relies on a survey question that requires a "yes" or "no" response, so it should not be biased by differences in national response styles. Also, unlike time spent helping children with homework, which depends on children's needs and the amount of homework that teachers assign, whether or not parents provide their children with a desk at home is a decision that is mainly up to parents. Further, unlike a measure such as the amount of time that parents spend talking with children about schoolwork, which appears to differ across racial/ethnic groups (Kao 2004), provision of a desk at home is a common way that parents can demonstrate the importance of education to their children in both East Asia and the U.S. Despite the utility of this indicator, the lack of other measures of parental involvement and data on parent-school relationships means that the major contributions of Chapter 4 are theoretical.

I have shown that among all high-achieving nations, the proportion of parents who had provided their fourth-graders with a desk at home is high—although not uniquely so—in Japan, around the international average in Taiwan and Singapore, and lower in Hong Kong than in the U.S. Specifically, the percentage of students with a study desk at home is 94 percent in Japan, 89 percent in Singapore, 87 percent in Taiwan, 77 percent in the U.S. and 72 percent in Hong Kong. This finding does not reflect a greater household income disparity in Hong Kong and the U.S. than in Japan, Taiwan and Singapore: while the average Gini coefficient of income inequality over the latter half of the 20th century as reported by Deininger and Squire (1996) is largest among all five nations in Hong Kong (41.6), it is significantly smaller in the U.S. (35.3) than in Singapore (40.1). [40]

Prevailing theory on the East Asian-U.S. parental involvement gap, which attributes the higher levels of parental involvement in East Asia to Confucianism, cannot explain why the percentage of parents who provide their children with a desk at home is lower in Taiwan than in Japan (the two nations in this study that are arguably most influenced by Confucian philosophy). Meanwhile, the U.S.-based theory on parental involvement that has received the most attention in recent sociological journals claims that parents' social-class background largely determines their involvement (Lareau 2000[1989], 2002, 2003). This theory cannot explain why 94 percent of parents in Japan provide their children with a desk at home. By drawing attention to the work of Coleman (1991, 1994) and Epstein (1986, 1995), who take into account the effect of the social organization of the school community on parents' decisions to become involved in their children's education, this dissertation has provided a theoretical way out of this morass.

Both Coleman (1991, 1994) and Epstein (1986, 1995) stress the role of social capital in the school community in influencing the proportion of parents who are involved in their children's education. Both Coleman and Epstein contend that school staff can work to develop

[40] According to Deininger and Squire (1996), the average Gini coefficient over the latter half of the 20th century in the U.S. (35.3) is just slightly larger than in Japan (34.8). It is smallest among all three nations in Taiwan (29.6).

the types of social capital that encourage parental involvement. Coleman would do this by creating information channels between parents and staff, while Epstein also emphasizes the importance of building relationships with parents that are based on mutual trust and respect. I build on these theories by taking into account the finding from social organizational research that concentrated disadvantage in a community tends to stymie the development of mutual trust and respect (Sampson, Raudenbush and Earls 1997). In so doing, I highlight how features of the school community influence the amount of work that school staff must undertake to develop the social capital needed to involve parents in the schooling process: staff at schools in communities of concentrated disadvantage are going to face more obstacles than are staff at schools in affluent communities. This argument leads us to expect that the likelihood of involvement varies across schools, even after controlling for parent background; that concentrated disadvantage is negatively associated with the likelihood of involvement, and that effective parent-school relationships are positively associated with parental involvement.

While the results are not conclusive, this dissertation provides evidence that is consistent with the preceding hypotheses. Two-level logistic regression models have shown that, holding constant school and student-level indicators of socioeconomic status, some of the variation in the likelihood of parental provision of a desk is due to differences between schools. The models have also shown that the socioeconomic status of the school community as well as the quality of parent-school relationships, as measured by school staffs' characterization of parental support, tend to be related to parental provision of a desk at home—even after taking parental characteristics into account. While these findings are correlations and thus do not necessarily indicate causation, they represent a step in the direction of gathering empirical evidence that is consistent with the argument that features of the school that children attend can influence their parents' decisions to become involved.

This dissertation has also developed two hypotheses to explain why the types of social capital necessary to promote parental involvement might be less plentiful in U.S. schools than in East Asian schools. First, I suggest that concentrated disadvantage in the school community limits the capacity of school staff to develop relationships with parents that are based on mutual trust and respect. This would explain why school staff in the U.S. are less able to encourage parental involvement if levels of concentrated disadvantage tend to be higher in the U.S. than in Japan, Singapore and Taiwan. Indeed, the analyses have demonstrated that the degree of concentrated disadvantage is higher among the lowest socioeconomic schools in Hong Kong and the U.S. than among those in Japan, Singapore and Taiwan. Given that the degree of income inequality in the U.S. is only slightly higher than in Japan, this finding is *not* simply reflecting the greater income inequality in the U.S. than in Japan, Singapore and Taiwan (Deininger and Squire 1996). Schools with the least affluent student populations in Hong Kong and the U.S. enroll a greater percentage of economically disadvantaged students compared to similar schools in Japan, Singapore and Taiwan. Unlike similar schools in Japan, Singapore and Taiwan, where the differences between the least affluent schools and other schools in terms of resources and neighborhood safety are not significant, the least affluent schools in Hong Kong and the U.S. are located in less safe neighborhoods than other schools in those nations (according to teachers' reports of feeling safe at school and in the surrounding area), and they have fewer resources that aid instruction than other schools. The political and institutional arrangements that allow for economic segregation of students in schools and for resources that aid instruction to be unequally distributed across economically segregated schools might be partially to blame for the lower proportion of involved parents in the U.S. compared to Japan, Singapore and Taiwan.

The other hypothesis to explain why the types of social capital conducive to fostering parental involvement might be less abundant in U.S. schools than in East Asian schools states that school staff in the U.S. are more apt to allow the socioeconomic status of the school community to influence their perceptions of parental involvement. I suggest that until school staff perceive of parental involvement as something that they can influence, as opposed to something that is determined by social class status, it is unlikely that they will invest their time and energy into encouraging parents to become involved in their children's education. The analyses have shown that the socioeconomic status of the school community and the degree to which teachers feel safe working in the school together account for a sizeable proportion of the variation in staff characterizations of parental involvement and support in the U.S. Additional evidence that school staff in the U.S. base their characterizations of parental support on social-class background comes from the fact that—in contrast to all East Asian nations—in the U.S., some of the relationship between staff characterizations of parental involvement and the proportion of parents at school who provide their children with a desk is spurious. These findings lend some support to the argument that teaching training and support programs in the U.S. have not done enough to prepare educators to involve parents in the schooling process.

IDEAS FOR FUTURE RESEARCH

Putting research on the East Asian-U.S. achievement gap on sounder empirical and methodological footing has shone the spotlight on the strengths and weaknesses in the literature and opened up several pathways for future research. The analyses of Chapter 2 lend support to

Stevenson and Stigler's (1992) argument that the achievement gap might stem from the relative uniformity of instruction across same-grade classrooms in East Asia compared to the U.S. Stevenson and Stigler point to the national curriculum standards of Japan and Taiwan to explain the relative uniformity of instruction there, and recommend that the U.S. adopt national curriculum standards to decrease the variability of instruction. While this dissertation did not examine the mechanisms through which educational systems encourage uniformity of instruction, it seems highly unlikely that national curriculum standards *on their own* are responsible for the degree of uniformity of instruction that Stevenson and Stigler witnessed in Japan and Taiwan. This is because only 7 of the 42 nations that participated in the 2003 or 2007 TIMSS did *not* have national curriculum standards. To understand how educational systems achieve significant uniformity of instruction across same-grade classrooms, future research might examine the relationship between uniformity of instruction and the various methods that nations use to monitor and enforce those standards.

The clarity and consistency of the standards themselves might also influence achievement—as well as the uniformity of instruction provided—regardless of the various methods used to hold schools and teachers accountable for the amount of instruction the provide If curriculum standards are easily accessible and transparent, then presumably a greater proportion of teachers will be aware of what they are supposed to teach and similarly, more parents will know what their children are supposed to be learning.

The analysis of Chapter 3 have shown that teachers in Japan and Taiwan do not vary much in their ability to promote math achievement, and that the amount of variation in teacher effectiveness within Hong Kong and Singapore is substantial. In so doing, this dissertation has opened up paths for future research on the structural features of educational systems that lead to more or less uniformity in teacher effectiveness. To understand which structural features contribute to greater uniformity in teacher effectiveness, future research might examine similarities and differences between Japan and Taiwan compared to Hong Kong and Singapore in teacher recruitment, training programs, hiring practices and professional support and development (for work in this direction see: Wang et al. 2003).

The implemented curriculum, or what students are taught, is clearly important. Previous work by Westbury (1992) contends that the reason East Asian math teachers appear to be more effective than their U.S. counterparts is because East Asian teachers cover more material in class than U.S. teachers typically cover. The underlying argument is similar to Sørensen and Morgan's (2000), which is that schools that teach more material produce more learning. Results from Chapter 3 do not support this position. Instead, the analysis has shown that among students who were taught the same number of math topics, average achievement scores are higher in East Asian nations than in the U.S. Future research might move beyond differences in quantity of instruction to focus on differences in the quality of instruction. For example, future research might focus on curriculum sequencing; pedagogical styles (for work in this direction see Stigler and Hiebert 1999; Leung 2005); textbooks and workbooks and the emphasis placed on theory versus practice in teacher training programs. Given that teachers in the U.S. tend to teach roughly the same percentage of TIMSS math topics as teachers in Singapore, future research might focus specifically on exploring the similarities and differences in the curriculum and pedagogy of those two nations (for work in this direction see Ginsburg et al. 2005).

Research on the relative academic success of East Asian students compared to their U.S. counterparts tends to underscore the role of East Asian parents in the achievement process. Immigrant studies as well as internationally comparative research contends that a greater proportion of East Asian parents are involved in their children's education, and more effectively involved, than U.S. parents. This dissertation has highlighted several empirical and theoretical weaknesses in this part of the literature. While previous research supplies some empirical evidence that a greater proportion of parents in East Asian nations than in the U.S. are involved in their children's education, there is little empirical support for the argument that East Asian parents are more effectively involved than U.S. parents because Confucianism places a high value on education. The only evidence that national cultural legacies lead to divergent parenting practices come from small-scale studies of immigrants to the U.S. (for example, Okagaki and Sternberg 1993). Future research might focus on collecting original data from comparable samples of parents in East Asian nations and the U.S. to better examine the relationship between beliefs in the importance of education and the nature of parental involvement (for similar research see: Yamamoto, Holloway and Suzuki 2006). Mickelson's (1990) work might inform this research. Mickelson contends that beliefs about the value of education are multidimensional. There are abstract beliefs, which reflect the societal consensus about the importance of education for future life chances, and concrete beliefs, which reflect the returns from education that students can expect to receive. Mickelson shows that concrete beliefs vary more among students than abstract beliefs and have more predictive power over student achievement than abstract beliefs. Future research might benefit by trying to measure parents' concrete beliefs about the value of education as opposed to their abstract beliefs.

Chapter 4 has emphasized the significance of the social resources that adhere in the relationships between parents and schools for increasing the proportion of involved parents in school communities. One of these is mutual trust and respect, which social organizational research has shown difficult to foster in communities of concentrated disadvantage (Sampson et al. 1997). I have shown that there are more school communities characterized by concentrated disadvantage in Hong Kong and the U.S. than in Japan, Singapore and Taiwan, which might explain why the proportion of involved parents is lower in the U.S. and Hong Kong than the other three nations. Future research might test this hypothesis by including more nations in the analysis.

Chapter 4 has shown that there is more concentrated disadvantage in school communities in Hong Kong and the U.S. than in Japan, Singapore and Taiwan. However, staff characterizations of parental involvement are less influenced by the socioeconomic status of the school community in Hong Kong than in the U.S. This is important because, as I have argued, school staff are less likely to devote time and energy into engaging parents who are less involved in their children's education if staff attribute parental involvement to social class as opposed to factors over which they have some control (e.g. parents' knowledge of their children's educational needs). U.S. education would thus benefit from understanding why staff at schools in communities of concentrated disadvantage in Hong Kong do not base their characterizations of parental support primarily on social class.

Although school staff can create and reinforce the social resources in the school community that encourage parental involvement, appropriate measures of these variables are lacking from internationally comparable datasets on early education. Future research on the mechanisms supporting higher proportions of involved parents in East Asian nations compared to the U.S. might gather high-quality data on the specific strategies school staff utilize to engage parents in the schooling process; the regularity with which they carry out these strategies, and the extent to which they attempt to reach out to less-involved parents. Documenting differences across educational systems in the extent to which teacher-training programs prepare teachers to engage parents in the schooling process might also help us understand system-level differences in the proportion of involved parents.

STRENGTHS AND WEAKNESSES IN THE DATA

In many ways, the quality of internationally comparable educational data circumscribe the research we are able to conduct. TIMSS has many strengths, which lead to greater insight into the impact of formal education on student achievement in math and science. This is for several reasons. First, TIMSS surveys 9-year-old students who are near the beginning of their formal education. This is beneficial because education is cumulative; for the majority of students surveyed, the elementary school they are attending is the only one they have ever attended. Therefore, the connections between the school inputs on the one hand, and student (or parent) outputs on the other are stronger than they are for students in their final year of secondary schooling. Additionally, unlike other "general use" tests that are used in international assessment, TIMSS strives to test the material that students learn in school. This practice allows us to examine more closely the relationship between the educational opportunities students receive and student achievement. Lastly, compared to most other international assessments, TIMSS provides tremendous insight into the curriculum of each educational system. The information gathered from the Teacher Questionnaire regarding the math topics on the TIMSS assessment that students have been taught makes TIMSS unique for allowing researchers to analyze the relationship between the quantity of instruction that students receive and the amount of knowledge students gained from that instruction.

There are some weaknesses in the TIMSS dataset, which limit the amount of insight research can gain into teacher effectiveness and the relationship between schools and parental involvement. Insight into teacher effectiveness is limited in part by the sampling design, which calls for participating educational systems to survey just one classroom per school. To distinguish the amount of variance in student achievement that is due to teachers from the amount that is due to schools, the survey needs to collect data from more than one fourth-grade classroom in each school. Fortunately, most educational systems collect data from more than one classroom in a school, but the numbers are small and thus make for less reliable samples. It is also difficult for research to gain traction on the factors that make some fourth-grade teachers more effective than others because TIMSS does not collect information on teaching methods and beliefs from fourth-grade teachers.

Although parents are one of the primary influences on student achievement in early elementary education, TIMSS does not offer much insight into the efforts educational systems, schools and teachers make to engage parents in the schooling process. The questions on the Principal Questionnaire regarding the school's expectations for parents do not do enough to differentiate schools by the importance they place on including parents. The survey instruments do not cover the strategies that school staff use to encourage parental involvement, nor do they ask teachers about the training or support they receive in inviting parents to play an active role in their children's education. It seems as though great progress on the reasons for achievement gaps between nations will not be made until international assessments take parents—and what schools do to engage parents—into account.

THE FUTURE OF DATA COLLECTION

To improve the quality of data gathered, and thus the quality of the work that scholars might conduct on international comparative education, I offer the following suggestions:

- *Change the default setting* Require that each school survey at least two, fourth-grade classrooms. This will provide better data with which to gauge teacher effectiveness within each educational system.
- *Ask about teaching methods and beliefs* Include on the fourth-grade Teacher Questionnaire sections concerning their teaching methods and beliefs about math. These are already included on the eighth-grade Teacher Questionnaire and could be easily adapted for the fourth-grade. The information garnered from these questions might bring us closer to understanding why some fourth-grade teachers are more effective than others at imparting math knowledge.
- *Make parents a priority* The Principal Questionnaire currently includes a section of questions on parental involvement that do tell us about the strategies that schools use to engage parents in the schooling process. Rather than focus on the school's expectations for parents, this section would ask about the communication that occurs between school staff and parents. For example, how frequently does the school communicate with parents? In what manner does the school typically communicate with parents? Does the school typically reach out to parents outside an official event only when a student is in trouble? Further, the Teacher Questionnaire would include at least one question about their training and/or preparation to include parents in the schooling process, or on the emphasis their school makes to communicate with parents. Ideally, the study would collect data from parents on their socioeconomic status and involvement methods, which the U.S. would make available for research.[41]

[41] The Progress in International Reading Literacy Study (PIRLS) already includes a Parent Questionnaire. However, there is no data from the U.S. on this portion of the study.

- *Survey more teachers* The greater the number of teachers surveyed about the school environment, the more accurate our understanding of the school environment becomes. Then, research might account for the bias and error in individual teacher's responses by constructing multilevel structural equations that correct for these shortcomings (Raudenbush and Sampson 1999a; Raudenbush and Sampson 1999b). Assuming that there are not many more fourth-grade teachers than those already included in the study, teachers in the third and fifth grade might be targeted. The questionnaire for these teachers would include not only background questions and questions pertaining to the school environment, but would also ask teachers the grade level at which they believe students should be taught each of the math topics that appear on the TIMSS test. This last section of questions would provide insight into the level of agreement among teachers regarding the math curriculum.
- *State participation increases* The U.S. participates in TIMSS and other international assessments as a single educational system. However, given that power over education rests with each state, it would make more sense if the individual states participated separately, like the Canadian provinces. The study could then create population weights such that data from each state might be combined to create a nationally representative sample.
- *Longitudinal studies* All current internationally comparative educational studies are cross-sectional, which means that scholars can examine relationships between variables but not cause. Conducting a longitudinal analysis would go a long way towards helping educators and policymakers understand the impact of educational inputs on student outputs.

Despite the areas of TIMSS (and internationally comparative educational datasets in general) in need of improvement, the chief conclusion of this dissertation is that these data can be used to conduct rigorous analyses.

References

Akiba, Motoko; Gerald K. LeTendre, and Jay P. Scribner. 2007. “Teacher Quality, Opportunity Gap, and National Achievement in 46 Countries.” *Educational Researcher* 36(7): 369-387.

Akiba, Motoko and Gerald K. LeTendre. 2009. *Improving Teacher Quality: the U.S. teaching force in global context.* New York: Teachers College Press.

Allison, Anne. 1996. “Producing Mothers.” pp. 135-155 in *Re-Imaging Japanese Women* edited by Anne E. Imamura. Berkeley, CA: University of California Press.

Barr, Rebecca and Robert Dreeban. 1983. *How Schools Work.* Chicago: University of Chicago Press.

Benjamin, Gail R. 1997. *Japanese Lessons: A Year in a Japanese School through the Eyes of an American Anthropologist and Her Children*. New York: New York University Press.

Carbonaro, William J. 2005. “Tracking, Students’ Effort, and Academic Achievement.” *Sociology of Education* 78(1): 27-49.

----. 1998. “A Little Help from My Friend’s Parents: Intergenerational Closure and Educational Outcomes.” *Sociology of Education* 71(4): 295-313.

Chen, Chuansheng and Harold W. Stevenson. 1989. “Homework: A Cross-cultural Examination.” *Child Development* 60(3): 551-561.

Cheung, Siu-Kau. 2001. “Parent Education Programmes in Hong Kong: Are They Effective?” *Hong Kong Journal of Social Work* 35(1): 85-96.

Chin, Tiffani and Meredith Phillips. 2004. “Social Reproduction and Child-rearing Practices: Social Class, Children’s Agency, and the Summer Activity Gap.” *Sociology of Education* 77(3): 185-210.

Chua, Amy. 2011. *Battle Hymn of the Tiger Mother*. New York: Penguin Press.

Chubb, John E. and Terry M. Moe. 1997. “Politics, Markets, and the Organization of Schools.” pp. 363-381 in *Education: Culture, Economy and Society*, edited by A. H. Halsey, H. Lauder, P. Brown and A. S. Wells. Oxford: Oxford University Press.

Coleman, James S.; Ernest Q. Campbell; Carol J. Hobson; James McPartland; Alexander M. Mood; Frederic D. Weinfeld and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

Coleman, James S. 1994. “A Vision for Sociology.” *Society* 32(1): 29-34.

----. 1991. *Parental Involvement in Education*, Policy Perspective Series. Washington, DC: Superintendent of Documents, U.S. Government Printing Office.

----. 1988. “Social Capital in the Creation of Human Capital.” *The American Journal of Sociology* 94: S95-S120.

Coleman, James S. and Thomas Hoffer. 1987. *Public and Private High Schools: The Impact of Communities.* New York: Basic Books, Inc.

Corcoran, Mary. 1995. “Rags to Rags: Poverty and Mobility in the United States.” *Annual Review of Sociology* 21: 237-267.

Darling-Hammond, Linda; Deborah J. Holtzman; Su Jin Gatlin, and Julian Vasquez Heilig. 2005. “Does Teacher Preparation Matter?: Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness.” *Education Policy Analysis Archives* 13(42): 1-48.

Dayal, Priyanka. 2012. “Most manufacturing jobs now require advanced skills.” Worcester *Telegram & Gazette*, Business Matters Cover story, March 4.

Deininger, Klaus and Lyn Squire. 1996. “A New Data Set Measuring Income Inequality.” *The World Bank Economic Review* 10(3): 565-91.

Desimone, Laura. 1999. “Linking Parent Involvement With Student Achievement: Do Race and Income Matter?” *The Journal of Educational Research* 93(1): 11-30.

Dika, Sandra L. and Kusum Singh. 2002. “Applications of Social Capital in Educational Literature: A Critical Synthesis.” *Review of Educational Research* 72(1): 31-60.

Dillon, Sam. 2010. “Top Test Scores From Shanghai Stun Educators.” New York *Times*, December 7, 2010. Last retrieved April 2, 2012: http://www.nytimes.com/2010/12/07/education/07education.html.

Domina, Thurston. 2005. “Leveling the Home Advantage: Assessing the Effectiveness of Parental Involvement in Elementary School.” *Sociology of Education* 78(3): 233-249.

Duhigg, Charles and Keith Bradsher. 2012. "How the U.S. Lost Out on iPhone Work." New York *Times* January 22.

Dumais, Susan A. 2006. "Early childhood cultural capital, parental habitus, and teachers' perceptions." *Poetics* 34(1): 83-107.

Epstein, Joyce L. and Mavis G. Sanders. 2000. "Connecting Home, School, and Community: New Directions for Social Research." pp. 285-306 in *Handbook of the Sociology of Education* edited by Maureen T. Hallinan. New York: Springer.

Epstein, Joyce L. 1996. "New Connections for Sociology and Education: Contributing to School Reform." *Sociology of Education* 69(Extra Issue on Sociology and Educational Policy: Bringing Scholarship and Practice Together): 6-23

----. 1995. "School/Family/Community Partnerships: Caring for the Children We Share." *The Phi Delta Kappan* 76(9): 701-712.

----. 1986. *Toward an Integrated Theory of School and Family Connection,* Report No. 3. Baltimore, MD: Center for Research on Elementary and Middle Schools.

Fan, Xitao and Michael Chen. 2001. "Parental Involvement and Students' Academic Achievement: A Meta-Analysis." *Educational Psychology Review* 13(1): 1-22.

Feliciano, Cynthia. 2005. "Educational Selectivity in U.S. Immigration: How Do Immigrants Compare to Those Left Behind?" *Demography* 42(1): 131-152.

Feuerstein, Abe. 2000. "School Characteristics and Parent Involvement: Influences on Participation in Children's Schools." *The Journal of Educational Research* 94(1): 29-39.

Flamm, Matthew. 2012. "Digital Jobs Central: How crazy is NYC's tech boom? Check out 568 Broadway." *Crain's New York business.com*. Retrieved online March 22, 2012: http://www.crainsnewyork.com/article/20120318/TECHNOLOGY/120319903

Foy, Pierre and Marc Joncas. 2004. "TIMSS 2003 Sampling Design." pp. 108-123 in *TIMSS 2003 Technical Report.* edited by M. O. Martin, I. V. S. Mullis and S. J. Chrostowski. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Gamoran, Adam. 1987. "The Stratification of High School Learning Opportunities." *Sociology of Education* 60(3): 135-155.

----. 1986. "Instructional and Institutional Effects of Ability Grouping." *Sociology of Education* 59(4): 185-198.

Ginsburg, Alan; Steven Leinwand; Terry Anstrom and Elizabeth Pollock. 2005. *What the United States Can Learn From Singapore's World-Class Mathematics System (and what Singapore* *can learn from the United States): An Exploratory Study.* U.S. Department of Education Policy and Program Studies Services.

Gonzalez, Eugenio J.; Joseph Galia; Alka Arora; Ebru Erberber, and Dana Diaconu. 2004. “Scaling Methods and Procedures for the TIMSS 2003 Mathematics and Science Scales.” pp. 252-273 in *TIMSS 2003 Technical Report.* edited by M. O. Martin, I. V. S. Mullis and S. J. Chrostowski. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Gordan, Robert; Thomas J. Kane and Douglas O. Staiger. 2006. “Identifying Effective Teachers Using Performance on the Job.” The Brookings Institute Discussion Paper 2006-01.

Hanushek, Eric A. 2011. “The Economic Value of Higher Teacher Quality.” *Economics of Education Review* 30(3): 466-479.

Hanushek, Eric A.; Dean T. Jamison; Eliot A. Jamison and Ludger Wößmann. 2008. “Education and Economic Growth: It’s not just going to school but learning something while there that matters.” *Education Next* 8(2): 62-70.

Hanushek, Eric A. and Steven G. Rivkin. 2010. “Generalizations about Using Value-Added Measures of Teacher Quality.” *American Economic Review: Papers & Proceedings* 100: 267-271.

Harding, David J. 2011. “Rethinking the Cultural Context of Schooling Decisions in Disadvantaged Neighborhood: From Deviant Subculture to Cultural Heterogeneity.” *Sociology of Education* 84(4): 322-339.

Harzing, Anne-Wil. 2006. “Response Styles in Cross-national Survey Research: A 26-country Study.” *International Journal of Cross Cultural Management* 6(2): 243-266.

Hoover-Dempsey, K. V.; J. M. T. Walker; H. M. Sandler; D. Whetsel; C. L. Green; A. S. Wilkins and K. E. Clossen. 2005. “Why do parents become involved? Research findings and implications.” *Elementary School Journal* 106(2), 105-130.

Hung, Chih-Lun. 2007. “Family, schools and Taiwanese children’s outcomes.” *Educational Research* 49(2): 115-125.

Joncas, Marc. 2004. “TIMSS 2003 Sampling Weights and Participation Rates.” pp. 186-223 in *TIMSS 2003 Technical Report.* edited by M. O. Martin, I. V. S. Mullis and S. J. Chrostowski. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

----. 2008. “TIMSS 2007 Sampling Weights and Participation Rates.” pp. 153-192 in *TIMSS 2007 Technical Report.* edited by J.F. Olson, M.O. Martin and I.V.S. Mullis. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Kane, Thomas J.; Jonah E. Rockoff and Douglas O. Staiger. 2006. “What Does Certification Tell Us About Teacher Effectiveness?: Evidence from New York City.” NBER Working Paper No. 12155. Cambridge, MA: National Bureau of Economic Research.

Kao, Grace. 2004. “Parental Influences on the Educational Outcomes of Immigrant Youth.” *International Migration Review* 38(2): 427-449.

----. 1995. “Asian Americans as Model Minorities? A Look at Their Academic Performance.” *American Journal of Education* 103(2): 121-159.

Kao, Grace and Jennifer S. Thompson. 2003. “Racial and ethnic stratification in educational achievement and attainment.” *Annual Review of Sociology* 29: 417-442.

Kerckhoff, Alan C. 1995. “Institutional Arrangements and Stratification Processes in Industrialized Societies.” *Annual Review of Sociology* 21:323-347.

Kuncel, Nathan R.; Marcus Credé and Lisa L. Thomas. 2005. “The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature.” *Review of Educational Research* 75(1): 63-82.

Lareau, Annette. 2003. *Unequal Childhoods: Class. Race. and Family Life*. Berkeley, CA: University of California Press.

----. 2002. “Invisible Inequality: Social Class and Childrearing in Black Families and White Families.” *American Sociological Review* 67(5): 747-776.

----. 2000[1989]. *Home Advantage: Social Class and Parental Intervention in Elementary Education,* 2nd ed. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Lareau, Annette and Erin McNamara Horvat. 1999. “Moments of Social Inclusion and Exclusion: Race, Class, and Cultural Capital in Family-School Relationships.” *Sociology of Education* 72(1): 37-53.

Lee, Valerie E. and David T. Burkam. 2003. “Dropping out of High School: The Role of School Organization and Structure.” *Educational Research Journal* 40(2): 353-393.

Lee, Shin-ying; Theresa Graham and Harold W. Stevenson. 1996. “Teachers and teaching: elementary schools in Japan and the United States.” pp 157-189 in *Teaching and learning in Japan* edited by Thomas P. Rohlen and Gerald K. LeTendre. New York: Cambridge University Press.

LeTendre, Gerald K. 1999. *Competitor or ally?: Japan’s role in American educational debates.* New York: Falmer Press.

Leung, Frederick K.S. 2005. "Some Characteristics of East Asian Mathematics Classrooms Based on Data from the TIMSS 1999 Video Study." *Educational Studies in Mathematics* 60(2): 199-215.

Lewis, Catherine C. 1995. *Educating Hearts and Minds: Reflections on Japanese Preschool and Elementary Education*. New York: Cambridge University Press.

Lockwood, J. R.; Daniel F. McCaffrey; Laura S. Hamilton; Brian Stecher; Vi-Nhuan Le and José Felipe Martinez. 2007. "The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures." *Journal of Educational Measurement* 44(1): 47-67.

Louie, Vivian. 2001. "Parents' aspirations and investment: The role of social class in the educational experiences of 1.5-and second-generation Chinese Americans." *Harvard Educational Review* 71:3. Retrieved October 25, 2007 http://www.edreview.org.ezp1.harvard.edu/harvard01/2001/fa01/f01louie.htm.

Ma, Liping. 1999. *Knowing and teaching elementary mathematics: teachers' understanding of fundamental mathematics in China and the United States.* Mahwah, NJ: Lawrence Erlbaum Associates.

Martineau, Joseph A. 2006. "Distorting Value Added: The Use of Longitudinal, Vertically Scaled Student Achievement Data for Growth-Based, Value-Added Accountability." *Journal of Educational and Behavioral Statistics* 31(1): 35-62.

Medrich, Elliot A. and Jeanne E. Griffith. 1992. *International Mathematics and Science Assessment: What Have We Learned?* National Center for Educational Statistics Research and Development Report 92-011. U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Last retrieved online April 28, 2012: http://nces.ed.gov/pubs92/92011.pdf.

Mickelson, Roslyn Arlin. 1990. "The Attitude-Achievement Paradox Among Black Adolescents." *Sociology of Education* 63(1): 44-61.

Montt, Guillermo. "Cross-national Differences in Educational Achievement Inequality." *Sociology of Education* 84(1): 49-68.

Morgan, Stephen L. and Aage B. Sørensen. 1999. "Parental Networks, Social Closure, and Mathematics Learning: A Test of Coleman's Social Capital Explanation of School Effects." *American Sociological Review* 64(5) 661-681.

Moynihan, Daniel P. 1965. *The Negro Family: The Case For National Action*. Washington, DC: Office of Policy Planning and Research, U.S. Department of Labor.

Mullis, Ina V.S.; Michael O. Martin; Graham J. Ruddock; Christime Y. O'Sullivan; Alka Arora and Ebru Erberber. 2005. *TIMSS 2007 Assessment Frameworks.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

The National Commission on Excellence in Education. 1983. *A Nation at Risk: The Imperative for Educational Reform.* Last retrieved on April 17, 2012 from: http://www2.ed.gov/pubs/NatAtRisk/index.html.

NCES. 2011. *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005-2009.* U.S. Department of Education. Last retrieved April 3, 2012: http://nces.ed.gov/nationsreportcard/pdf/studies/2011458.pdf.

Nye, Barbara; Spyros Konstantopoulos and Larry V. Hedges. 2004. "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26(3): 237-257.

OECD. 2005. *Teachers Matter: Attracting, developing and retaining effective teachers.* Paris, France: OECD Publishing.

Okagaki, Lynn and Peter A. Frensch. 1998. "Parenting and Children's School Achievement: A Multiethnic Perspective." *American Educational Research Journal* 35(1): 123-144.

Palardy, Gregory J. and Russell W. Rumberger. 2008. "Teacher effectiveness in the first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, 30(2): 111-140.

Peak, Lois. 1991. *Learning to Go to School in Japan: The Transition from Home to Preschool Life*. Berkeley, CA: University of California Press.

Perry, Michelle. 2000. "Explanations of Mathematical Concepts in Japanese, Chinese, and U.S. First and Fifth-Grade Classrooms." *Cognition and Instruction* 18(2): 181-207.

Pierce, Reuben G. 1993. "The Learning Gap: Why Our Schools Are Failing and What We Can Learn from Japanese and Chinese Education by Harold W. Stevenson and James W. Stigler." *The Journal of Negro Education* 62(3): 394-5.

Pong, Suet-ling; Lingxin Hao and Erica Gardner. 2005. "The Role of Parenting Styles and Social Capital in the School Performance of Immigrant Asian and Hispanic Adolescents." *Social Science Quarterly* 86(4): 928-950.

Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24: 1-24.

Rabe-Hesketh, Sophia and Anders Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata, Second Edition.* College Station, TX: A Stata Press Publication.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition.* Thousand Oaks, CA: Sage Publications.

Raudenbush, Stephen W.; Brian Rowan and Sang Jin Kang. 1991. "A Multilevel, Multivariate Model for Studying School Climate with Estimation Via the EM Algorithm and Application to U.S. High-School Data." *Journal of Educational Statistics* 16(4): 295-330.

Reardon, Sean F. and Stephen W. Raudenbush. 2009. "Assumptions of Value-Added Models for Estimating School Effects." *Educational Finance and Policy* 4(4): 492-519.

Rivkin, Steven G.; Eric A. Hanushek; and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73(2): 417-458.

Roksa, Josipa and Daniel Potter. 2011. "Parenting and Academic Achievement: Intergenerational Transmission of Educational Advantage." *Sociology of Education* 84(4): 281-298.

Rothstein, Richard. 2004. *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap.* Washington, DC: Economic Policy Institute.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.

Sampson, Robert J. and Stephen W. Raudenbush. 1999a. "Econometrics: Towards a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods." *Sociological Methodology* 29(1): 1-41.

----. 1999b. "Assessing Direct and Indirect Effects in Multilevel Designs With Latent Variables." *Sociological Methods & Research* 28(2): 123-152.

Sampson, Robert J.; Stephen W. Raudenbush and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277(5328): 918-924.

Sandefur, Gary D., Ann M. Meier and Mary E. Campbell. 2006. "Family Resources, social capital, and college attendance." *Social Science Research* 35(2): 525-553.

Schneider, Barbara. 2000. "Social Systems and Norms: A Coleman Approach." pp. 365-385 in *Handbook of the Sociology of Education* edited by Maureen T. Hallinan. New York: Springer.

Schneider, Barbara and Yongsook Lee. 1990. "A Model for Academic Success: The School and Home Environment of East Asian Students." *Anthropology & Education Quarterly* 21(4): 358-377.

Singapore Ministry of Education. 2012. “The Public as Partners.” Singapore: Ministry of Education. Last retrieved on February 21, 2012 from: http://www.moe.gov.sg/about/#our-mission.

Sørensen, Aage B. and Maureen T. Hallinan. 1977. “A Reconceptualization of School Effects.” *Sociology of Education* 50(4): 273-289.

Sørensen. Aage B. and Stephen L. Morgan. 2000. “School Effects: Theoretical and Methodological Issues.” pp. 137-160 in *Handbook of the Sociology of Education* edited by Maureen T. Hallinan. New York: Springer.

Stanton-Salazar, Ricardo D. 1997. “A Social Capital Framework for Understanding the Socialization of Racial Minority Children and Youths.” *Harvard Educational Review* 67(1): 1-41.

Stanton-Salazar, Ricardo D. and Sanford M. Dornbusch. 1995. “Social Capital and the Reproduction of Inequality: Information Networks among Mexican-Origin High School Students.” *Sociology of Education* 68(2): 116-135.

Stevenson, David L. and David P. Baker. 1991. “State Control of the Curriculum and Classroom Instruction.” *Sociology of Education* 64(1): 1-10.

Stevenson, Harold W.; Shin-ying Lee and James W. Stigler. 1986. “Mathematics Achievement of Chinese, Japanese, and American Children.” *Science* 231(4739): 693-699.

Stevenson, Harold W. and James W. Stigler. 1992. *The Learning Gap: Why our Schools are Failing and What We Can Learn from Japanese and Chinese Education.* New York: Summit Books.

Stevenson, Harold W.; Chuansheng Chen and Shin-Ying Lee. 1993. “Mathematics Achievement of Chinese, Japanese, and American Children: Ten Years Later.” *Science* 259(5091): 53-58.

Stigler, James W. and James Hiebert. 1999. *The Teaching Gap: Best Ideas from the World’s Teachers for Improving Education in the Classroom.* New York: The Free Press.

Sun, Yongmin. 1998. “The Academic Success of East-Asian-American Students—An Investment Model.” *Social Science Research* 27(4): 42-456.

Teachman, Jay D.; Kathleen Paasch and Karen Carver. 1996. “Social Capital and Dropping out of School Early.” *Journal of Marriage and the Family* 58(3): 773-783.

Wang, Aubrey H.; Ashaki B. Coleman; Richard J. Coley; Richard P. Phelps. 2003. *Preparing Teachers Around the World. Policy Information Report.* Princeton, NJ: Educational Testing Service.

Westbury, Ian and Chao-Sheng Hsu. 1996. "Does State Control of the Curriculum Matter? A Response to Stevenson and Baker." *Educational Evaluation and Policy Analysis* 18(4): 343-347.

Westbury, Ian. 1992. "Comparing American and Japanese Achievement: Is the United States Really a Low Achiever?" *Educational Researcher* 21(5): 18-24.

U.S. Department of Education. 2009. "Race to the Top Program. Executive Summary." Washington, DC: U.S. Department of Education.

Van de Werfhorst, Herman G. and Jonathan J.B. Mijs. 2010. "Achievement Inequality and the Institutional Structure of Educational Systems: A Comparative Perspective." *Annual Review of Sociology* 36: 407-428.

Vinovskis, Maris A. 1999. *The Road to Charlottesville: The 1989 Education Summit.* Washington, D.C: National Education Goals Panel. Last retrieved April 2, 2012: http://govinfo.library.unt.edu/negp/reports/negp30.pdf.

Yamamoto, Yoko; Susan D. Holloway, and Sawako Suzuki. 2006. "Maternal involvement in preschool children's education in Japan: Relation to parenting beliefs and socioeconomic status." *Early Childhood Research Quarterly* 21(3): 332-346.

Yano, Hirotoshi. 1993. "What Can We Learn from the Learning Gap?" *Educational Researcher* 22(1): 36-7+43.