Educational Evaluation and Policy Analysis
Spring 2001, Vol. 23, No. 1, pp. 1–17

Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement

David P. Baker
Motoko Akiba
Gerald K. LeTendre
Alexander W. Wiseman
Pennsylvania State University

The growth of structured, outside-school activities for improving students' mathematics achievement is an enduring feature of modern schooling with major policy implications. These "shadow education" activities mimic, or shadow, formal schooling processes and requirements. Using extensive cross-national data from the Third International Mathematics and Science Study, we examine shadow education as a macro-phenomenon of modern schooling through its (a) prevalence, (b) strategies for use, and (c) associated national characteristics. We find that shadow education is prevalent worldwide, but that there is considerable cross-national variation in its use. Contrary to findings from single country studies, we find most shadow education is remedial in nature. We then test hypotheses concerning the national origins of shadow education and its impact on nations' production of mathematics achievement. Our results show that institutional factors of education, including limited access and lower levels of funding, drive the use of shadow education, instead of high-stakes testing and national achievement incentives. We conclude by discussing implications for both educational policy and theory regarding the degree to which institutionalization of mass schooling increasingly dominates contexts of schooling.

The relationship of outside-school learning with higher levels of achievement across students has been a recurring theme in policy discussions arising from findings of cross-national achievement differences in mathematics during the 1980s and 1990s (Bray, 1999; McDonough, 1997; Stevenson & Baker, 1992; Stevenson, Schiller, & Schneider, 1994; Stevenson & Stigler, 1992). Of interest to education policymakers are national consequences resulting from large-scale use of structured, supervised, outside-school learning in the form of tutoring, review sessions, proprietary cram schools, and related practices in order to increase students' mastery of academic subjects in school. The impact of recent international comparisons of mathematics achievement production focuses attention on studies highlighting the extensive use of these outside-school activities in top achieving countries such as Japan and Korea (e.g., Stevenson & Stigler, 1992). The interjection of outside-school learning into the debate over cross-national achievement makes these activities, and the reasons large numbers of students undertake them, central to discussions about how countries can improve their national production levels of mathematics and science achievement (e.g., National Research Council, 1999). What has not been well understood is the origin of these growing outside-school activities as a macro-phenomenon of modern schooling.

Related to this policy topic, the global use of outside-school learning activities represents a deeper theoretical issue concerning the incorporation of schooling into the social structure of modern societies throughout the world. As formal education becomes (a) the primary formal institution by which social status is channeled across genera-

tions within families, (b) the major generator of a society's technical knowledge (and human capital), (c) a major socializing agent to an increasingly complex adult world, and (d) a recognized institutional partner in the formation of citizenry for the modern nation state, its role in everyday life takes on greater intensity and import for families and students (Meyer, 1977). Given this, the development of extensive outside-school activities aimed at within-school achievement becomes an important and enduring feature of modern schooling. Researchers studying the dynamic relationship between school and family influences on achievement have come to refer to the host of structured outside school achievement activities as *shadow education* (Stevenson & Baker, 1992).

The term *shadow education* conveys the image of outside-school learning activities paralleling features of formal schooling used by students to increase their own educational opportunities (see also Bray, 1999; George, 1992; LeTendre, 1994; Tsukada, 1991).[1] These activities go well beyond routinely assigned homework; instead they are organized, structured learning opportunities that take on school-like processes. The after-hours cram schools found in some Asian counties, such as *juku* in Japan, are the most extreme in mimicking in-school forms. But there are a wide variety of activities that share a similar logic such as correspondence courses, one-on-one private tutoring, examination preparatory courses, and full-scale preparatory examination schools (e.g., Japanese *yobiko*). For example, systems of tutoring are extensive in Hong Kong, Singapore, Taiwan (*Bu-shyi-ban)*, Korea (*Hakwon),* Greece, and Turkey.

Observers of education systems throughout the world have commented on the growth in shadow educational activities (Bray, 1999; de Silva, 1994; Foondun, 1998; Hussein, 1987; Kwan-Terry, 1991; Stevenson & Baker, 1992).[2] But beyond descriptive reports about shadow education from one country to the next, all of the empirical research on these activities has been limited to assessments of their impact on the educational achievement of students, or on related outcomes such as educational inequality across family socioeconomic status (SES) (e.g., Stevenson & Baker, 1992). As our analysis of data from 40 countries finds here, although shadow education is a widespread phenomenon, there is considerable cross-national variation in both the level of participation in shadow education among students and the intended goals of these activities. Understanding which factors propel certain countries to develop extensive shadow education aimed at specific educational goals is crucial for education policy development that includes this dimension of modern schooling, as well as for a deeper theoretical understanding of the phenomenon. We suggest several hypotheses concerning the origin, persistence, and characteristics of shadow education (a) because of the prevalence and emphases of cross-national achievement comparisons and educational policies relying on analyses of factors contributing to variations in achievement and (b) because many argue that shadow education contributes to cross-national variation in achievement.

The Third International Mathematics and Science Study (TIMSS) provides rare data on individual student use of shadow education from nationally representative samples of a relatively large and educationally diverse set of nations, enabling us to address these issues. Three sets of analyses are undertaken with these data. The first set aims at answering the question of how prevalent shadow education is cross-nationally. Most of the literature focuses on just a few spectacular cases of national shadow education without any real analysis of the use of these activities across a wide spectrum of national school systems. The TIMSS data allow educational policymakers for the first time to see the degree to which these activities are used across numerous and varied national educational systems.

The second set of analyses examines cross-national variation in the primary role (remedial versus enhancement) that shadow education plays in countries. Although most discussion of the logic behind shadow education focuses on a predominant enrichment strategy, there are ample reasons to suspect that there are widespread remedial strategies behind its use by students and their families. It also is possible that across populations of students within nations there is a mixture of strategies propelling the use of shadow education. An international examination of trends in the goals of students in their use of shadow education has never been done before.

The third set of analyses addresses three hypotheses about the origins of shadow education, plus one final hypothesis about shadow education's ultimate impact on national levels of achievement. Because the theoretical arguments and hypotheses behind these analyses are complex, we describe each separately.

Hypotheses on Cross-National Origins of Shadow Education

What attributes of educational systems produce extensive use of shadow education across large proportions of students and for what purposes? Further, what do these explanations reveal about the structure of formal schooling and the production of mathematics knowledge across nations?

These kinds of questions require answers at the national level. We are interested here neither in why particular students use these mechanisms, nor which characteristics of students are associated with use.[3] The theoretical arguments tested consider cross-national variation in institutional factors that link formal schooling with shadow education. If shadow education activities parallel formal school processes, with the latter providing the former their structure, value, and eventual impact on educational outcomes, understanding exactly which formal school processes activate and expand outside-school activities is a key to understanding the growth of these outside-school activities.

There are a number of ideas about which characteristics of formal schooling shadow education responds to; many of which originate from different views on the causes and correlates of cross-national achievement differences. The most popular argument is that large-scale shadow education is a product of a widely practiced, highly publicized *enrichment strategy*. This argument suggests that extensive shadow education exists in systems in which there are intense competitions for future educational opportunities accompanied by "tight linkages" between academic performance and later life opportunities in the labor market.[4] In short, this argument is that shadow education will be observed as a system-wide enrichment strategy when there are clear, high-stakes decision points within the system that are based on performance tests. For example, it is often argued that qualities such as (a) the tight connection between elite universities and excellent labor market opportunities in Taiwan, (b) the use of secondary school certificate examinations in the labor market in Hong Kong, (c) the secondary school selection process in Japan, and (d) the examination-based link between secondary schools and the best universities in Greece produce a strong logic for students and families to use extensive shadow education (Katsillis & Rubinson, 1990; Lin, 1983; Mitchell, 1968; Stevenson & Baker, 1992; Sweeting, 1983). All of these examples revolve around a high-stakes test that is a public gatekeeper to education and labor market opportunities.

This argument suggests the following hypothesis about cross-national variation in the use of shadow education:

> (H1) The use of high-stakes tests in national educational systems results in the prevalence of shadow education activities as an enrichment strategy across students.

A contrasting argument, and one that has not received much attention in the debates over national achievement, is that shadow education is a product of a widespread, institutionalized *remedial strategy*. The notion here is that schooling and one's ability to move through it with at least a minimal level of academic success has become such a central and crucial task for such a wide portion of the youth population, that prematurely ending a school career has taken on extremely severe social consequences for both the individual and society as a whole. Over the past century, the widespread use of mass, compulsory schooling with large public investments and enhanced standing as a social sector has made schooling the central, formal institution connecting children and youth to adult status (e.g., Fuller & Rubinson, 1992). The central message of schooling is that (a) all must participate and at least minimally advance and achieve when and where they can, and (b) there are few, if any, traditional routes to adult status left in most modern societies (e.g., status inheritance, sinecures, or blatantly ascriptive status attainment). Not only are individuals faced with the inevitability of mounting a long formal school career, but also state policymakers, for a range of issues from economic development to health and social welfare, continually advance educational solutions to reduce most societal problems. This institutional process makes categories such as school dropouts and low achievers prime targets for public policy, which aims at educational solutions to a host of social problems. This frequently leads to calls for yet more schooling, prolonged educational participation across cohorts of youth, and greater scholastic achievement for formal education's ameliorating effects on society.

This central institutional quality of formal schooling is so pervasive and has such a "taken-for-granted" nature to it that in comparison to more salient characteristics, such as high-stakes tests, it is often passed over when thinking about how school dynamics influence students and families. It is also clear that although a mass, compulsory pub-

lic school system has become a dominant model worldwide, there is still cross-national variation in the degree to which nations embrace this model (e.g., Baker & Holsinger, 1996). Given this, it is reasonable to argue that extensive shadow education will be an outgrowth of a more fully implemented model of these qualities of modern schooling within a nation.

This argument suggests the following hypothesis about cross-national variation in shadow education:

> (H2) In nations with more pronounced qualities of mass schooling, such as high expenditures on public education or large enrollment rates at both elementary and secondary levels, shadow education will be more prevalent as a remedial strategy.

A related, but counter argument is that national economic development will be negatively related to the use of shadow education because of the greater use of high-stakes tests, especially the use of high-stakes tests as a type of hyper-process of educational qualification in developing nations, resulting from an often hypothesized process in developing systems known as the "diploma disease" (Dore, 1976). The argument is that less economically developed nations will rely more on high-stakes testing and qualification-earning, and that this will create a prevalence of shadow education as an enrichment strategy in these nations (i.e., H1 will be confirmed, but H2 will be in the opposite direction hypothesized).[5]

A third argument on the generation of shadow education revolves around the notion that some nations (e.g., Korea, Singapore, Hong Kong, Japan) have developed systemwide motivation for national achievement in subjects like mathematics and science through a variety of means, and a high degree of shadow education is a consequence of an explicit *national strategy* to foster national achievement. As the development of human capital takes on qualities and responsibilities of a unilateral national mission for successful global competition, educational policies aimed at achievement can be formed from a specific nationalist perspective promoted by the state. This argument suggests that these perspectives can spread through both the faculty and students within a system and provide an enhanced motivation for national achievement (Benavot, Cha, Kamens, Meyer, & Wong, 1991). In the literature on cross-national differences in schooling and achievement this argument has been expanded to include cultural scripts for high persistence on academic tasks (Stevenson, Lummis, Lee, & Stigler, 1990). The notion is that some nations produce a more salient set of cultural and national motivations for students to persist in their efforts towards academic achievement, including taking publicly recognized international tests. Extending this to outside school activities leads to the argument that more shadow education activities of either a remedial or enrichment purpose will be generated in nations with greater emphases on student persistence and achievement. This reasoning leads to the following hypothesis about cross-national variation in shadow education:

> (H3) In nations with a higher collective emphasis on achievement in mathematics, as reflected in average student persistence in finishing national tests, shadow education (remedial or enrichment) will be more prevalent.

In addition to these predictions about which factors generate shadow education across national systems, there are similar arguments about the possible impact of widespread shadow education on national achievement. The main argument is that the presence of extensive shadow education will be positively associated with national achievement levels. The logic behind this idea is twofold. First, extensive shadow education represents extended academic training for a large proportion of the student population and hence has the potential to increase overall achievement. Second, shadow education is a consequence of larger national processes that encourage or require intensive efforts in subjects like mathematics through such devices as high-stakes testing or other tightly linked accountability mechanisms in these systems (Bishop, 1998; National Research Council, 1999). These suggest two related hypotheses about shadow education and national achievement in mathematics:

> (H4a) Nations with more extensive shadow education use among students will have higher achievement means;
>
> (H4b) and a substantial portion of any impact of high-stakes tests on national achievement will work indirectly through shadow education.

Methods

Data

The data are from the 1994–95 TIMSS and include large samples of seventh- and eighth- grade stu-

dents from 41 countries.[6] A two-stage sampling design was used in TIMSS. First, a probability proportional to the size of the school sample was selected from a pool of all schools in each participating country with eligible students in the target age-year cohort. Second, up to two mathematics classrooms per school were sampled using an equal probability of selection sampling design. Sampling weight takes into account effects of stratification or disproportional sampling of subgroups and makes appropriate adjustments for non-response. The achievement data used in our analyses are from seventh- and eighth-grade achievement tests in mathematics.[7] We also used information provided by background questionnaires completed by students and school questionnaires completed by principals.

Dependent Variable Measures

We based our indicators of national shadow education use and prevalence on the following TIMSS student questionnaire item: "*During the week, how much time before or after school do you usually spend taking extra lessons/cram school in mathematics?*" The original response options range from no time to more than 5 hours a week. Close translations of this question from the original nation-specific questionnaires in Chinese, Japanese, and Spanish were conducted, and the closely translated meaning of the item was found to capture the full meaning of shadow education in each national system and each country's language. Further analysis of each country's original wording confirmed that the item was appropriately comparable across nations.

Using responses to this item, we created six national indicators of shadow education. Each captures unique dimensions of participation in these activities. The first three indicate national-level use of shadow education. The last three indicate predominant educational strategies driving national levels of participation.

1. *Percentage using shadow education.* This is the most global participation variable and is the percentage of all students indicating any participation in shadow education activities.

2. *Mean hours of participation in shadow education.* This indicates the level of use among students and is the national mean of hours per week spent participating in shadow education activities.

3. *Intensity of shadow education use.* We multiplied the first two variables together (percentage participating x mean hours of participation) to create an indicator of the intensity of shadow education use in the nation. Specifically, this indicates the level of concentration and emphasis on shadow education activities in national education systems.[8]

4. *Percentage use among low math achievers.* To determine whether student achievement characteristics correspond with shadow education activity participation we calculated the percentage of shadow education participation by students producing the lowest one third of a given country's math scores.

5. *Percentage use among high math achievers.* Likewise, we calculated the percentage of shadow education activity participation by students producing the highest one third of a given country's math scores.

6. *Modal strategy of use (remedial vs. enrichment).* This indicator is the logit regression coefficient of use of shadow education (use vs. no use) as a function of math performance.

Since the TIMSS questionnaire did not ask students why they used shadow education, we estimated their reasons by examining the nature of the relationship between shadow education and math performance at the student level in each country. We did this by regressing a dichotomous indicator of students' weekly shadow education use on their math score (b6). We also included a set of control variables previously shown to influence decisions to use shadow education (Stevenson & Baker, 1992). Since it is known that the economic and educational status of families, and, in some cultures, the gender of the student, may influence the tendency to use shadow education, there are four control variables for students and their families: a composite measure of socioeconomic status (b1)[9], whether or not the home language was different from the language of instruction at school (b2), sex (b3), and type of community in which the family resides (b4). Similarly, extra-lesson remedial resources (b5) provided directly by schools could influence the use of shadow education, so we included this school variable also.

Using these predictors, the modal strategy of use is estimated by a logit procedure for each nation's seventh- and eighth-grade sample:[10]

Shadow education use = b1(SES) + b2(home language) + b3(sex) + b4(community type) + b5(at-school remedial) + b6(math score) + b7(SES × math score) + a1

Independent Variable Measures

To further explore the hypotheses described above, we included several indicators of national-level educational development as well as system characteristics that account for structural elements of mass compulsory schooling as indicated by our hypotheses.

1. *High-stakes testing.* We took this variable of curriculum-based external exit exam prevalence from Bishop (1998) and coded it into 0 for no use of high-stakes testing, irregular or regional use of high-stakes testing, or use only in math ($N = 22$), and into 1 for the regular use of high-stakes testing both in math and science ($N = 19$).[11]

2. *Public expenditure as percentage of GNP.* We obtained this variable of public expenditure as a percentage of GNP in 1994 from United Nations Educational, Scientific, and Cultural Organization (UNESCO) Statistical Yearbooks (1990, 1991, 1992, 1993, 1994). We supplemented the absence of data for the Russian Federation in 1993 with data from the UNESCO *Statistical Yearbook* (1997). This variable ranges from 1.34% in Hong Kong to 5.26% in Norway. The mean is 3.51%.

3. *Gross enrollment ratio.* We used the total gross enrollment ratio at elementary and secondary levels in 1995 from the UNESCO *Statistical Yearbook* (1997). For some countries whose enrollment ratio was unavailable in 1995, we used the available ratio in adjacent years.

4. *Mean student response rate.* We used the percentage of seventh- and eighth-grade students completing the math achievement test in each nation as a measure of national motivation and persistence in math as a cultural factor. We used this measure because, as argued above, Stevenson (1991) and Stevenson et al. (1990) have found that student task persistence and motivation are often results of national scripts for high student achievement. The lowest response rate was 87.0% for French-speaking Belgium and the highest response rate was 99.7% for Japan. The mean student response rate was 95.7%.

Results

I. How Prevalent is Shadow Education Cross-Nationally?

As shown in Figure 1, the use of shadow education by eighth graders to study mathematics is extensive throughout the TIMSS sample of 41 counties.[12] The cross-national average suggests that more than a third of all seventh- and eighth-graders weekly participate in tutoring sessions, cram schools, or other forms of shadow education activity. Furthermore, of the entire international sample of students represented here, 4 out of 10 (39.6%) regularly participate in shadow education activities to improve their mathematics achievement in school. Other analyses of the hours of weekly participation not shown in this figure indicate that while most weekly participation consists of 1 hour or less, one fifth of the entire TIMSS student sample undertook shadow education activities of 2 or more hours per week.

Also, a sizable standard deviation for percent using shadow education (about 21 percentage points) provides clear evidence of substantial cross-national variation in the use of shadow education among seventh- and eighth-graders. More than three fourths of the seventh- and eighth-graders in nations like Japan, South Africa, Philippines, and Colombia weekly participate in shadow education activities, while in countries like Denmark, England, Norway, and Germany less than 20% participate.

II. What is the Primary Role of Shadow Education in Most Nations?

The TIMSS student questionnaire did not directly ask students why they participate in shadow education activities, but we can infer possible reasons from a series of bivariate and multivariate analyses of other information provided by the study. The key component is the relationship between students' mathematics test scores and shadow education participation across all students. Nations where this relationship is clearly positive indicate that an enrichment strategy is the dominant role of shadow education—that is, students with high performance in mathematics tend to use shadow education for strategic advantages in future educational contests. Nations where this relationship is clearly negative indicate that a remedial strategy is the dominant role of shadow education—that is, students with low performance in mathematics tend to use shadow

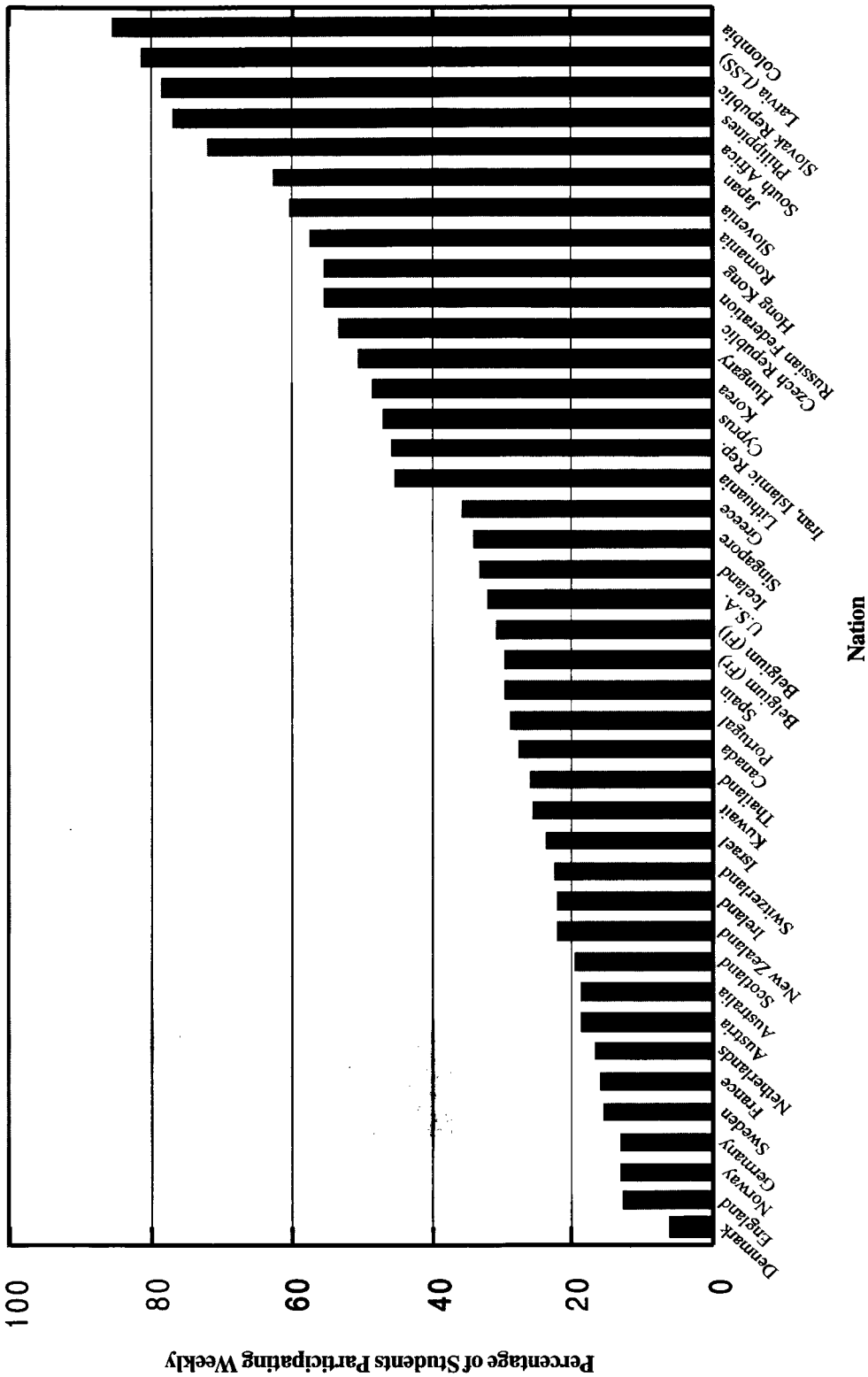

FIGURE 1. *Shadow education participation by nation, eight-grade mathematics.* Note: *Includes any amount of participation.* M = *37.6%,* SD = *21.3%,* N = *41.*

education to maintain minimal or otherwise acceptable achievement levels in school.

We first examined the bivariate relationship between shadow education participation and math ability for each country. Shadow education in each country is classified as predominately remedial if substantially more low math ability students than high math ability students participate.[13] This classification of shadow education strategies does not mean that in a system we classify as *remedial* there is no enrichment motivation; rather, *remedial* captures the modal trend of use among that nation's students. For example, using this method we can classify nations such as Cyprus, Israel, Belgium, and Denmark as modally remedial because, compared to students who scored in the top one third, two to over three times as many students using shadow education produced the lowest one third of math scores.[14] In other countries the remedial strategy is less dominant, but it is still substantial. For example, among students who use shadow education in the U.S., Germany, and Kuwait, approximately 30% more score low on math than score high. After confirming these bivariate patterns with the multivariate analysis described below, we classified these countries and others with similar participation trends as remedial for the purpose of hypothesis testing. The role of shadow education in a country is classified as predominantly enrichment if substantially more high math scoring students use it. For example, in Korea and Romania almost twice as many high scoring students regularly use shadow education than do low scoring students (42% vs. 24% and 40% vs. 27%, respectively).

In a few countries the relationship between math scores and participation was too weak to classify as either remedial or enrichment because virtually equal amounts of high and low scoring students regularly participated in shadow education activities. For example, in Japan and Hong Kong about a third of both low and high scoring students use shadow education. We classified these countries as *mixed* (i.e., remedial and enrichment strategies in similar proportions).

To further refine and verify our classifications of the modal patterns of shadow education use, we estimated the modal strategy behind shadow educational participation using a multivariate procedure. In general, these results confirm our classifications from the bivariate relationship.[15] Figure 2 graphs the unstandardized partial regression coefficient, which indicates the association between students' math scores and their shadow education participation. These coefficients summarize the degree to which math scores increase participation. Accordingly, the direction of the coefficient indicates for all students whether participation is more remedial or enrichment-oriented after controlling for other factors known to be associated with individual decisions to participate in shadow education.

Summarizing across these 41 separate equations, we find it notable that in over three fourths (31) of the countries the modal strategy for using shadow education is for remedial purposes (i.e., negative, statistically significant coefficients for the relationship between student achievement score and shadow education use). For example, the probability of participation in shadow education activities increases by 3.5% for each one-point decrease in TIMSS mathematics score in countries like Denmark, Germany, and the U.S. In only 3 of the 41 countries in the study is shadow education predominately an enrichment strategy (i.e., positive, statistically significant coefficients). For example, for each one-point increase on the mathematics test, Korean students were almost 7% more likely to participate in shadow education. In the remaining eight countries there is a substantial mixture of both strategies (i.e., statistically non-significant, or very small coefficients). Of particular interest is our finding that nations such as Greece, Hong Kong, and Japan, which have been widely cited as producing widespread enrichment-oriented shadow education, show significant levels of remedial use.[16]

III. Which National Factors Influence the Origin, Prevalence, and Character of Shadow Education?

Table 1 displays the results from a series of equations that tested hypotheses concerning the association between various national-level educational development and system factors and six indicators of national use of shadow education. The first column tests H1, the second two columns test H2, and the last column tests H3.

Across all of the indicators of shadow education, the presence of high-stakes tests in a country is unrelated to the use of shadow education (H1). High-stakes tests neither systematically engage a larger proportion of a nation's student population in the use of shadow education, increase time spent on shadow education, nor increase the intensity of

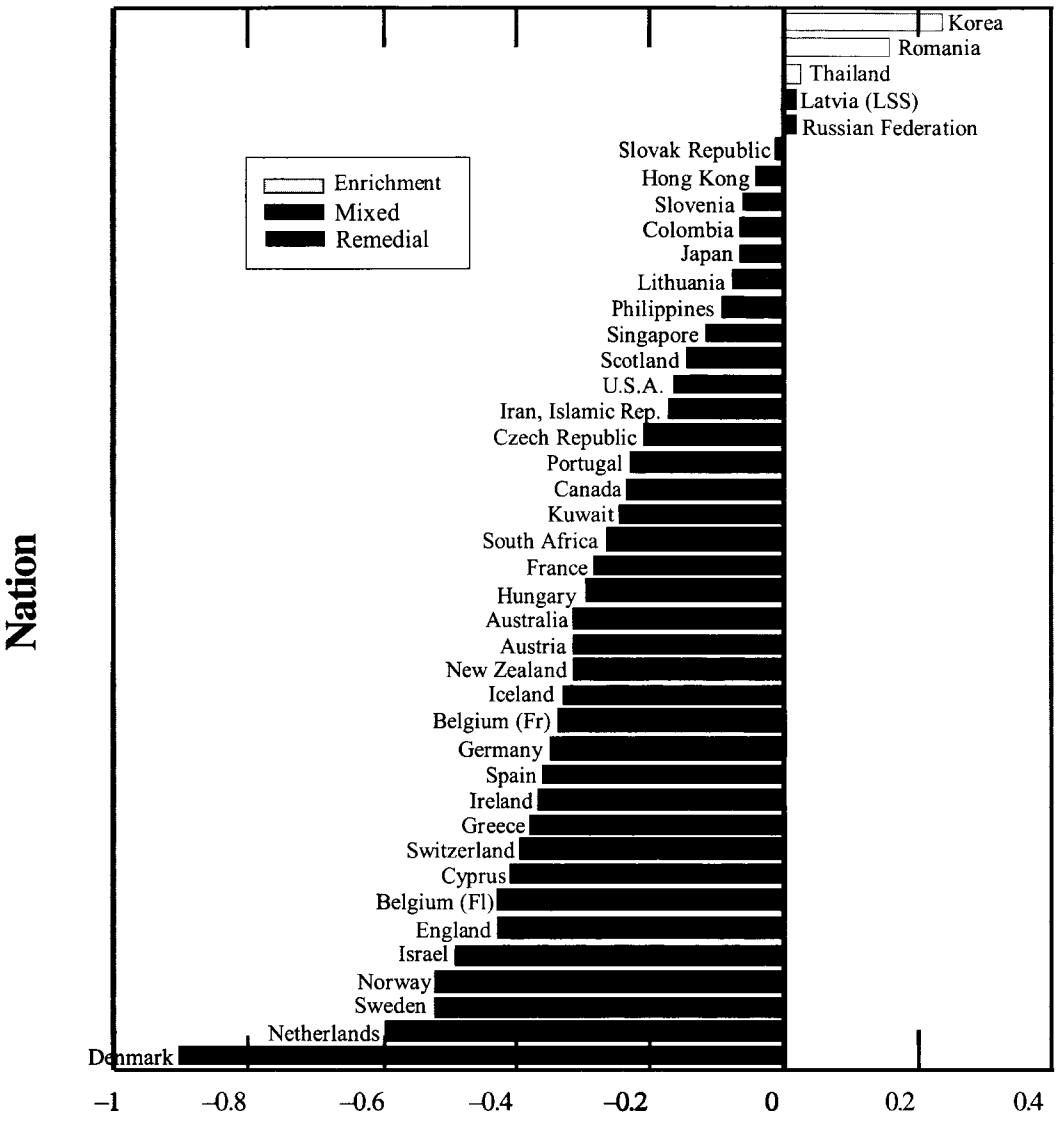

FIGURE 2. *Classification of national modal strategy motivating shadow education.* Note: *Logit Coefficient of effects of math score on shadow education participation, controlling for SES, home language, sex, community, remedial teaching, and the interaction term between SES and math score.* Enrichment *equals significant* ($p < .01$) *positive coefficient;* Mixed *equals non-significant coefficient;* Remedial *equals significant* ($p < .01$) *negative coefficient.*

use of shadow education. Similarly, high-stakes tests are neither associated with any particular national modal strategy of use nor use by low or high mathematics achievers. In short, our results indicate that high-stakes tests are not related at all to cross-national variation in the use of shadow education.

Contrary to this, we find evidence that the degree to which mass education is institutionalized within a nation is related to national use of shadow education. In all six equations, either or both indicators of mass education—public expenditures and gross enrollment rates—are related to various measures of national shadow education. And, taken together, these two factors account for 25% to 40% of the cross-national variation in various indicators of national use of shadow education.[17]

TABLE 1

OLS Estimates of the Influence of Institutional Qualities of National School Systems on Indicators of National Use of Shadow Education (N = 41)

Dependent variables (Indicators of national shadow education use)	Independent variables (Institutional qualities of schooling)				R^2
	High-stakes testing	Public expenditure as % of GNP	Gross enrollment rate	Mean student response rate	
Percent using shadow education	1.17 (6.04)	−8.27* (3.43)	−0.47* (.22)	−0.78 (1.19)	0.35
Mean hours of participation in shadow education	−1.91[02] (.76)	−1.57[02] (.04)	−7.99[03]** (.003)	9.20[03] (0.2)	0.23
Intensity of shadow education use[a]	1.74 (1.024)	−13.64* (5.81)	−.98* (.379)	−.66 (2.0)	0.38
Use among low math achievers[c]	58 (2.05)	4.78** (1.17)	.27** (.08)	−.72 (.40)	0.53
Use among high math achievers[d]	.75 (1.87)	−2.85* (1.06)	−.24** (.07)	.45 (.37)	0.45
Modal strategy (remedial vs. enrichment)[b]	4.49[03] (.06)	−.10** (.03)	−6.96[03]** (.002)	1.83[02] (.01)	0.44

[a] Calculated by multiplying national percentage using shadow education × mean hours of participation.

[b] Logit coefficient from national student-level equation predicting shadow education participation as a function of students' math ability.

[c] Percentage of participating students using shadow education among the lowest one third of math scores in each country.

[d] Percentage of participating students using shadow education among the highest one third of math scores in each country.

*$p < .05$. **$p < .01$.

Surprisingly, these effects are generally not in the predicted direction. Contrary to the logic of the second hypothesis (H2), more pronounced qualities of mass education do not generate a higher prevalence of shadow education use, rather shadow education becomes a private supplement to a less than fully developed mass education system. Both lower public educational expenditures and less than full enrollments at the seventh and eighth grades lead to more use of shadow education, greater intensity of use, and more reliance on remedial strategies as well as use by high achieving students. Only among low achieving students does more mass incorporation of education produce more shadow education as specifically predicted by H2.

Lastly, there is no evidence that high national levels of motivation to complete international tests are associated with greater use of shadow education (H3). The mean student response rate is not associated with any of the six indicators of shadow education.

IV. Does Shadow Education Influence the Production of National Mathematics Achievement?

The first three hypotheses and analyses determined the origin, prevalence, and characteristics of shadow education. The final analysis considers the fourth hypothesis that national use of shadow education will be associated with national achievement levels. To test this, we estimated six separate equations regressing national mean mathematics scores on the six indicators of national use of shadow education as well as the presence of high-stakes tests. The Ordinary Least Squares (OLS) estimates for each equation are shown in Table 2, and it is clear that across all six indicators there is no influence of national shadow education on national levels of mathematics achievement.[18] In analyses not reported here, we also found that interaction terms between each indicator of shadow education and high-stakes tests were also not related to national achievement levels.[19]

Conclusion

What do all of these results indicate about shadow education cross-nationally and its role in the production of national achievement? First, by the middle of the 1990s when the TIMSS data was collected, large numbers of seventh and eighth graders all over the world regularly used some form of structured outside-school instruction aimed at improving their mathematics performance in school. Even though the modern school now provides extensive instruction over an increasingly expanding school day and growing school year in several nations, many students and their families further add to that through a host of shadow education activities.

Second and surprisingly, amidst this overall trend of greater use of structured outside learning activities, the primary role that shadow education plays in the average national system represented in the TIMSS sample is that of remedial training: Low performing math students turn to these activities at much greater rates than high performing students. Although there is some mix of both remedial and enrichment strategies in every national sample, evidence of remedial uses outweighs enrichment uses in most nations. The image, developed in prior literature, of shadow education as a widely used enrichment strategy is not well supported by these data.

Third, although there is a global tendency toward remediation, there is evidence of considerable cross-national variation in the use and modal goals of shadow education across this sample of nations. The use, and intensity of use, of these outside-school learning activities varies from country to country. There are also some distinguishable differences in overall amounts of remedial and enrichment strategies behind the use of shadow education. Among the more dominant pattern of remedial use, some countries' modal use is clearly for enrichment purposes, while others almost equally mix remedial and enrichment shadow education among students.

Fourth, the prediction that high-stakes tests or national motivations to perform well on international tests would increase the presence of shadow education did not receive support. Although evidence that lower public expenditures increases national levels of shadow education is in line with the general idea of national economic development and the "diploma disease," the fact that the other crucial part of this argument is not supported, namely that high-stakes tests are not associated with shadow education, casts some doubt on the notion that developing nations do something different in educational development than developed nations (see Heyneman, 1987, for further discussion of testing in developing nations).

TABLE 2
OLS Estimate of Influence of Indicators of National Shadow Education Use and High-Stakes Tests on National Seventh- and Eighth-Grade Mathematics Achievement (N = *41*)

Dependent variable	Independent variables		R^2
	Indicators of shadow education	High-stakes test	
National math mean	Percent using shadow education −.5 (.41)[a]	24.2 (16.8)	0.08
National math mean	Mean hours participation in shadow education 33.2 (35.5)	19.6 (16.9)	0.06
National math mean	Intensity of shadow education use[b] −.2 (.24)	23.6 (17.0)	0.06
National math mean	Modal strategy (remedial vs. enrichment)[c] 1.1 (39.1)	21.4 (17.3)	0.04
National math mean	Use among low math achievers[d] .2 (1.0)	21.9 (17.1)	0.04
National math mean	Use among high math achievers[e] −.5 (1.2)	23.2 (17.4)	0.05

[a] OLS coefficient (standard error).
[b] Calculated by multiplying national percentage using shadow education × mean hours of participation.
[c] Logit coefficient from national student-level equation predicting shadow education participation as a function of students' math ability.
[d] Percentage of participating students using shadow education and producing the lowest one third of math scores in each country.
[e] Percentage of participating students using shadow education and producing the highest one third of math scores in each country.

Fifth, there is clear evidence that shadow education is part of the institutional character of mass education. Although the results are not in the precise direction originally hypothesized, there is an association between (a) both levels of expenditures on education and enrollment rates and (b) the degree to which shadow education is present in national systems, which suggests important implications for educational policies.

Lastly, our results indicate no association between national levels of shadow education use and national mathematics achievement levels. Although these activities have been shown to increase some students' performances, and even though in many nations large numbers of students use them, they do not appear to have an overall influence on raising national levels of achievement.

Implications for Educational Policy

These results hold two major implications for the education policymaker. First, these outside-school structured learning activities are a major undertaking in many nations. For a variety of reasons, families and students continue to augment formal schooling through the investment of private resources in their children's mathematics education. This is the first time that there is systematic, comparable data on the cross-national use of shadow education, so it is difficult to tell if this is a waning or waxing trend overall. It is clear, however, that this is a widespread education practice in all types of national systems. This means that for many students in many nations, outside-school learning represents far more than assigned homework and informal help from parents. Educators must appreciate the dimensions of shadow education and its potential to impact upon formal instruction and management of students and curricula (see also Bray, 1999). The long-term effects of large-scale augmentation to formal instruction are often hard to predict, but they can have dramatic impacts. For example, other research has shown that within systems that heavily use

shadow education, the differential ability of families to invest in large amounts of shadow education can exacerbate achievement outcomes by family socioeconomic status and can even lead to curricular and programmatic crises in public education (Stevenson & Baker, 1992).

A related example is the case of the extensive system of shadow education in Japan and its unintended influence on formal school policies. For example, shadow education in this country plays an integral role in providing feedback to families and students concerning changes in admission to secondary schools and universities. Also, some Japanese educational reform attempts have been stalled because of reliance on shadow education to perform institutional tasks and outside-school learning has been shown to reduce the work of Japanese public school teachers through standardizing many students' achievement levels in classes (LeTendre, 1994; 1996). These functions, which unofficially serve formal systems and families, can have an impact on educational policies.

Second, although these activities are prevalent and potentially a source of influence on formal schooling, the education policymaker should understand that they do not represent a simple route to improving national levels of mathematics production. The notion that tighter links between education and work opportunities, controlled by high-stakes tests, leads to motivation for greater outside-school learning activities, which in turn raise nations' average levels of achievement, is not supported by these data. This is not to say that high-stakes tests (or the more general policy reform idea now very active in the U.S. and some other nations of aligning curriculum standards, assessment, and opportunity) do not have national consequences for the improvement of achievement production in schools. Rather the data here indicate that this process, to the degree that it drives national achievement, does not seem to work through the extensive networks of shadow education operating in many national systems. It should be noted that since all of these results are from seventh and eighth grade, tight-linkages, high-stakes tests, and shadow education might be more closely associated cross-nationally at later points in schooling, such as at the end of secondary school. Shadow education probably has some effects on individual students, but either the effects are not intensive enough to change overall national levels of achievement, or the effects are relatively even throughout most of the TIMSS sample of nations.[20] The latter, however, is less likely since, as is shown here, there is considerable cross-national variation in level of participation in shadow education.

Implications for Educational Theory

These data illustrate the general notion that national and cross-national patterns of shadow education lead to an important institutional perspective on modern schooling. The widespread prevalence of shadow education and intensity of its use in schooling processes of students and families dramatically underscores the degree to which education increasingly dominates the structure of childhood and child rearing. Neo-institutional theorists have argued that social institutions profoundly construct the way life is organized (e.g., DiMaggio & Powell, 1983). That argument applied to formal schooling (e.g., Meyer, Tyack, Nagel, & Gordon, 1979) is well supported by these cross-national data on shadow education. As schooling becomes the primary institution for the generation of knowledge and its transmission to generations of children in the form of achievement, one may expect the continued growth in outside-school activities that are specially aimed at children's performance within schools. Homework and the assumption of extensive parental involvement in academics are now common features of the formal educational institution; shadow education takes this same process a step further.

Although our hypothesis that cross-national variation in qualities of mass education would be related to shadow education is counter to the actual direction of the association found in the data, the results nonetheless illustrate the general point about institutionalization. In constructing this hypothesis prior to the analysis, we were unaware of the overwhelming remedial use of shadow education cross-nationally. We assumed, as is so often described in case studies of shadow education, that a larger portion of nations would show a dominant enrichment pattern as a way for families to buy advantage in an already well-funded system. This proved to not be the case; the data indicate that underfunded and underenrolled national systems of education produce more shadow education, and a large portion of this is remedial (this may also happen in underfunded parts of overall well-funded national systems, as in the U.S.). As education expands and intensifies as an institution, families in less funded and less enrolled (probably represent-

ing issues of access) systems use shadow education. Shadow education augments the formal public system, and this can be a very extensive augmentation.

One way to think of this is that mass schooling sets the stage for the increasing importance of education as an institution (e.g., Meyer et al., 1979; Meyer, Ramirez, & Soysal, 1992) and, to the degree that this process creates greater demand for quality schooling than is supplied, augmentation through shadow education is likely. Keep in mind, too, that no country lacked some measurable level of shadow education, and that many economically well-funded education systems also have extensive shadow education. As mass schooling as a world norm continues to intensify the importance of schooling, shadow education itself becomes an institutionalized component of mass education that grows and expands. In some countries it has been observed that, like formal schooling, shadow education activities even take on non-academic rationales for participation. TIMSS case study data from Japan, for example, shows adolescents participating in shadow education because their peers do, and it is a vehicle for socializing as well as for study (OERI, 1998).

A second way to consider these findings is that lower levels of funding and enrollment may mean lower confidence in available school quality and hence, in some nations, shadow education develops as a kind of market reaction to underdevelopment. This has tended to be overlooked since shadow education was initially identified and most intensely studied in Japan as an enrichment strategy (Tsukada, 1981; Rohlen, 1980, Stevenson & Baker, 1992; OERI, 1998). But Japan's system of rich mixes of both remedial and enhancement shadow education is less common and grew out of unique historical conditions of an under-supply of academic track high schools in the 1970s coupled with the inability of the government to reform an entrance examination system to a highly hierarchical set of Japanese universities (Schoppa, 1991; Roesgaard, 1998). Before the evidence presented here on the substantial remedial nature of shadow education in many nations, shadow education was assumed to be a market reaction to intensified competition for further educational and labor market opportunities—a sort of market reaction to overdevelopment. Clearly, both processes, market reactions to intense educational development and underdevelopment, are evident among these nations. Further, these results give a greater appreciation for the capacity of underdeveloped educational systems to produce large demand for shadow education. Moreover, the complexity of these trends suggests that institutional theory about the dynamic growth of education phenomena in modern society needs further conceptual development.

Theories on how educational structures change in modern society have not considered the potential for mass public schooling to create the conditions needed to generate high levels of shadow education as is documented here. The key to developing such a theoretical understanding is to see how mass schooling, greater legitimization of academic achievement, and a heightened sense of demand for quality schooling come together in systems to produce large-scale shadow education. As shadow education loosens the boundaries between public control of education and private educational activities, there are both theoretical and policy lessons to be learned. Although private educational activities predate the creation of modern public schooling, mass public schooling ironically produces a logic and demand for mass shadow (i.e., private) education in many countries. Furthermore, education policymakers should appreciate the ability of widespread shadow education to add a significant non-public component to the process without the production of full-scale private schools. This has the potential to change the governance structure of education by introducing private services to a wider host of families and students, hence merging families and private educational resources into new institutional arrangements.

Notes

[1] *Shadow education* as a term is also used to describe informal adult education activities provided by firms internally versus externally through formal education programs (e.g., Craig & Evers, 1981).

[2] See Bray (1999) for a thorough description of this phenomenon worldwide.

[3] Although student-level information helps us characterize national patterns of use later in this paper.

[4] "Tight linkage" in this literature has come to mean the perception of a strong relationship between a student's academic performance, usually on a major examination, and later labor market opportunities.

[5] See Dore (1997) and Little (1997) for more recent nuances of the original argument.

[6] The TIMSS study used a strict review of sampling procedures in each participating nation to ensure basic technical standards. In U.S. government reports on

TIMSS, even stricter sampling standards were noted past the data collection. Although in U.S. reports some nations' results are noted, all participating nations' samples met overall basic TIMSS requirements (see Gonzalez & Smith, 1997).

[7] Using an Item Response Theory (IRT) scoring procedure with randomly rotated items on the achievement test and an imputation process, five plausible values of math achievement for each student were estimated by the TIMSS study. Extensive preliminary analyses show that the substantive results remained the same for all the plausible values. Therefore, the computation results from just the first plausible value are reported here.

[8] Some may argue that if, in a given country, there is a high amount of shadow education used by a small population, then our calculation of intensity would be misleading because it would suggest a low intensity of shadow education throughout that nation when actually a high degree of shadow education is used by some of the population. However, in a previous study of the variation in shadow education (Akiba & LeTendre, 1999) we found that few nations had high levels of participation by small segments of the population. Shadow education tends to be diffused at relatively low levels (less than 2 hours per week) in most systems. Only in South Africa, Slovenia, Korea, and Singapore did we find more than 10% of students taking more than 3 hours of shadow education per week. Therefore, we constructed our variable as a general estimate of the amount of usage considering both the population size using shadow education and the mean hours of usage in the nation as a whole.

[9] Families' socioeconomic status includes parents' education, possessions at home, and measures of cultural capital (e.g., student visits museums, attends concerts, etc.).

[10] A logit estimation procedure is used since the dependent variable is categorical and varies substantially from 50/50 in some countries. The equation is estimated as

$$P = \exp(L_1)/[1 + \exp(L_1)] - \exp(L_0)/[1 + \exp(L_0)]$$

where P is the probability of participation; L_0 is the logit before the unit change in X_i; and L_1 is $L_0 + B_i$, the logit after the unit change in X_i.

[11] High-stakes testing variable (Bishop, 1998) is used here because it is based on the categorization of the nations in TIMSS. Bishop reports that this variable is coded using information from comparative education studies and education encyclopedias and from the embassies of TIMSS nations.

[12] Three are only 40 countries here, but studies like TIMSS have traditionally divided Belgium into Flemish- and French-speaking due to significant educational differences.

[13] *Substantial* is defined here as a statistically significant difference of a magnitude of at least 30 percentage points. In analyses not reported here this classification was further verified by the relationship between math scores and the amount of time spent in shadow education.

[14] The percentages of low performing students using shadow education in Cyprus, Israel, Belgium, and Denmark are 40, 47, 44, and 68.5, respectively.

[15] For all of the analyses below, we classified countries as *mixed* in the few cases that our multivariate analysis found very small, yet statistically significant, relationships between participation and math scores.

[16] On average, the percentage of students taking shadow education was highest in the nations with *mixed* modal (64%), followed by the nations with *enrichment* modal (43.8%) and the nations with *remedial* modal (27.2%) (see Akiba & LeTendre, 1999).

[17] The cross-national variation was obtained from the comparison of the coefficient of variations (R^2) between the model with public expenditures and gross enrollment rates and the model without these variables. Results are not reported here but are available from the authors.

[18] Although it is not of major theoretical concern here, our analysis also found no direct effect of high-stakes tests on national achievement. See also Bishop's (1998) report, which reports modest evidence of an effect. Further, we also added the public expenditure in education as a percentage of GNP variable to each of the equations; there was no change in the non-significance of shadow education *or* high-stakes testing after controlling for expenditure. Analyses are available from the authors upon request.

[19] Two other variables of high-stakes testing were also examined for their associations with the national math mean in the six models. The first variable was coded as 0 for no use of high-stakes testing, 1 for use only in math or science, or 2 for use in both math and science. The second variable was coded as 0 for no use, 1 for use at least in math or science. All of these variables were not significantly related to the national math mean.

[20] This is also not to say that these activities do not have measurable effects on individual students' improvement in mathematics. There is some evidence to suggest that they do have a varied marginal impact (Bray, 1999; Stevenson & Baker, 1992).

References

Akiba, M., & LeTendre, G. (1999, April). *Remedial, mixed and enhancement systems: An analysis of math achievement and extra-lessons.* Paper presented at the Comparative International Education Society, Toronto, Canada.

Baker, D. P., & Holsinger, D. B. (1996). Human capital formation and school expansion in Asia: Does a unique regional model exist? *International Journal of Comparative Sociology, 37*(1–2), 159–173.

Benavot, A., Cha, Y., Kamens, D., Meyer, J. W., & Wong, S. (1991). Knowledge for the masses: World models and national curricula, 1920–1986. *American Socio-*

logical Review, 56(1), 85–100.

Bishop, J. (1998). *Do curriculum-based external exit exam systems enhance student achievement?* (CPRE Research Report Series RR-40). Philadelphia: Consortium for Policy Research in Education, Graduate School of Education, University of Pennsylvania.

Bray, M. (1999). *The shadow education system: Private tutoring and its implications for planners.* Paris: UNESCO: International Institute for Educational Planning.

Craig, R. L., & Evers, C. J. (1981). Employers as educators: The shadow education system. *New Directions for Experiential Learning; Business and Higher Education: Toward New Alliances, 13*(13), 29–46.

de Silva, W. A. (1994). *Extra-school tutoring in the Asian context with special reference to Sri Lanka.* Maharagama: Department of Educational Research, National Institute of Education.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48,* 147–160.

Dore, R. (1976). *The diploma disease. Education, qualification and development.* University of California Press: Berkeley, CA.

Dore, R. (1997). *Japan, internationalism and the UN.* New York: Routledge.

Foondun, A. R. (1998). *Private tuition: A comparison of tutoring practice in Mauritius and some Southeast Asian countries.* Bangkok: UNICEF East Asia & Pacific Regional Office.

Fuller, B., & Rubinson, R. (Eds.). (1992). *The political construction of education: The state, school expansion, and economic change.* Praeger: New York.

George, C. (1992, April 4). Time to come out of the shadows. *Straits Times* (Singapore).

Gonzalez, E. J., & Smith, T. A. (Eds.). (1997). *User guide for the TIMSS international database: Primary and middle school years 1995 assessment.* Chestnut Hill, MA: Boston College.

Heyneman, S. P. (1987). Uses of examinations in developing countries: Selection, research, and education sector management. *International Journal of Educational Development, 7*(4), 251–263.

Hussein, M. G. A. (1987). Private tutoring: A hidden educational problem. *Educational Studies in Mathematics, 18*(1), 91–96.

Katsillis, J., & Rubinson, R. (1990). Cultural capital, student achievement, and educational reproduction: The case of Greece. *American Sociological Review, 55*(2), 270–279.

Kwan-Terry, A. (1991). The economics of language in Singapore: Students' use of extracurricular language lessons. *Journal of Asian Pacific Communication, 2*(1), 69–89.

LeTendre, G. K. (1994). Distribution tables and private tests: The failure of middle school reform in Japan. *International Journal of Educational Reform, 3*(2), 126–136.

LeTendre, G. K. (1996). Constructed aspirations: Decision-making processes in Japanese educational selection. *Sociology of Education, 69*(3), 193–216.

Lin, C. (1983). The Republic of China (Taiwan). In R. M. Thomas & T. W. Postlethwaite (Eds.), *Schooling in East Asia: Forces of change* (pp. 104–135). New York: Pergamon.

Little, A. (Ed.). (1997). The diploma disease twenty years on [Special issue]. *Assessment in Education: Principles, Policy and Practice, 4*(1).

McDonough, P. M. (1997). Access, equity, and the privatization of college counseling. *Review of Higher Education, 20*(3), 297–317.

Meyer, J. W. (1977). The effects of education as an institution. *American Journal of Sociology, 83*(1), 55–77.

Meyer, J. W., Ramirez, F. O., & Soysal, Y. N. (1992). World expansion of mass education, 1870–1980. *Sociology of Education, 65*(2), 128–149.

Meyer, J. W., Tyack, D., Nagel, J., & Gordon, A. (1979). Public education as nation-building in America: Enrollments and bureaucratization in the American states, 1870–1930. *American Journal of Sociology, 85*(3), 591–613.

Mitchell, R. (1968). *Pupil, parent and school: A Hong Kong study.* Taipei: Orient Cultural Service.

National Research Council. (1999). *High Stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press.

Office of Educational Research and Improvement. (1998). *The educational system in Japan: Case study findings.* Washington, DC: U.S. Department of Education.

Roesgaard, M. (1998). *Moving mountains.* Aarhus, Denmark: Aarhus University Press.

Rohlen, T. (1980). The *Juku* phenomenon: An exploratory essay. *Journal of Japanese Studies, 6*(2), 207–242.

Schoppa, L. (1991). *Education reform in Japan: A case of immobilist policies.* New York: Routledge.

Stevenson, D. L., & Baker, D. P. (1992). Shadow education and allocation in formal schooling: Transition to university in Japan. *American Journal of Sociology, 97*(6), 1639–1657.

Stevenson, D. L., Schiller, K. S., & Schneider, B. (1994). Sequences of opportunities for learning. *Sociology of Education, 67*(3), 184–198.

Stevenson, H. (1991). Japanese elementary school education. *The Elementary School Journal, 92,* 109–20.

Stevenson, H. W., Lummis, M., Lee, S., & Stigler, J. W. (1990). *Making the grade in mathematics: Elemen-*

tary school mathematics in the United States, Taiwan and Japan. Reston, VA: National Council of Teachers of Mathematics.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. Summit Books: New York.

Sweeting, A. (1983). Hong Kong. In R. M. Thomas & T. W. Postlethwaite (Eds.), *Schooling in East Asia: Forces of change* (pp. 272–297). New York: Pergamon.

Tsukada, M. (1991). *Yobiko life: A study of the legitimation process of social stratification in Japan*. Berkeley: University of California–Berkeley.

United Nations Educational, Scientific, and Cultural Organization. (1990). *Statistical yearbook*. Paris: Author.

United Nations Educational, Scientific, and Cultural Organization. (1991). *Statistical yearbook*. Paris: Author.

United Nations Educational, Scientific, and Cultural Organization. (1992). *Statistical yearbook*. Paris: Author.

United Nations Educational, Scientific, and Cultural Organization. (1993). *Statistical yearbook*. Paris: Author.

United Nations Educational, Scientific, and Cultural Organization. (1994). *Statistical yearbook*. Paris: Author.

United Nations Educational, Scientific, and Cultural Organization. (1997). *Statistical yearbook*. Paris: Author.

Authors

DAVID P. BAKER is Professor of Education and Sociology at Pennsylvania State University, Educational Theory & Policy Program, 300 Rackley Building, University Park, PA 16802; dpb4@psu.edu. He specializes in sociology of education and comparative education policy analysis.

MOTOKO AKIBA recently completed her Ph.D. in the Education Theory and Policy Program at Pennsylvania State University (mailing address above); mxa238@psu.edu. Her areas of specialization are comparative education and school violence.

GERALD K. LETENDRE is Associate Professor of Education at Pennsylvania State Univiersity (mailing address above); aww112@psu.edu. He specializes in comparative education and policy analysis.

ALEXANDER W. WISEMAN is a Ph.D. candidate in the education Theory and Policy Program at Pennsylvania State University (mailing address above); aww112@psu.edu. His areas of specialization are comparative education and school management.

Manuscript received April 6, 2000
Revision received September 29, 2000
Accepted October 25, 2000