

# Measurement Issues in Research on Shadow Education: Challenges and Pitfalls Encountered in TIMSS and PISA

MARK BRAY AND MAGDA NUTSA KOBAKHIDZE

Expanding numbers of researchers are focusing on the scale and impact of private supplementary tutoring. Such tutoring is widely called shadow education, since much of its curriculum mimics that of regular schooling. Although shadow education has expanded significantly worldwide and is now recognized to have far-reaching significance, research faces methodological and conceptual challenges. This article focuses on analyses of shadow education data from the Third (or Trends in) International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). An initial problem arises from definitions of shadow education and therefore from research focus. Further challenges arise from the initial phrasing and then translation of items in international questionnaires. The article notes that some studies have been grounded in problematic data, which has led to misleading pictures. Methods and approaches are maturing, but much refinement remains necessary for an adequate understanding of the nature and implications of shadow education.

The research literature on private supplementary tutoring has expanded significantly since the turn of the century.<sup>1</sup> This partly reflects the increased importance of the phenomenon. Private tutoring is a major global phenomenon that demands considerable expenditure, consumes a great deal of the time of children and their families, provides extensive employment, and has a backwash on the operation of regular schooling. The research literature has helped to show the nature of the phenomenon, including the amounts and types of tutoring received by different grades of students in a wide range of countries. Parts of the literature have moved beyond simple statistics on the proportions of students receiving tutoring and have identified the content of that tutoring, the intensity in hours per week, and seasonal variations at different times of year. Other dimensions include the psychological and social impact of tutoring and relationships between out-of-school tutoring and in-school learning.

Within the literature, private supplementary tutoring is widely called shadow education.<sup>2</sup> This metaphor has been adopted because much of the

<sup>1</sup> See, e.g., Kwok (2004), Lee et al. (2009), Mori and Baker (2010), Aurini et al. (2013), and Bray et al. (2013).

<sup>2</sup> See, e.g., Bray (1999), Buchmann (2002), Lee et al. (2009), and Aslam and Atherton (2012).

Received September 7, 2013; revised May 9, 2014; accepted June 12, 2014; electronically published October 8, 2014

*Comparative Education Review*, vol. 58, no. 4.

© 2014 by the Comparative and International Education Society. All rights reserved.  
0010-4086/2014/5804-0004\$10.00

content of private supplementary tutoring mimics that of regular schooling. As the curriculum in the regular schools changes, it is suggested, so does the curriculum in the shadow; and as the regular schooling sector expands, so does the shadow. This metaphor is useful, though may also be restrictive. Thus, some forms of private tutoring supplement the regular curriculum in the sense of providing additional content rather than mimicked content. Issues of vocabulary and focus need attention from a methodological perspective since they can contribute to confusion. This has been evident in both large-scale quantitative studies and small-scale qualitative ones.

The focus of this article is on two large-scale quantitative studies, both of which have served multiple countries and education systems. One is called TIMSS, initially the Third International Mathematics and Science Study (1995 and 1999) and later the Trends in International Mathematics and Science Study (since 2003).<sup>3</sup> It is organized under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). The other major cross-national study is called PISA (Programme for International Student Assessment) and is organized under the auspices of the Organisation for Economic Co-operation and Development (OECD). This article shows that although these studies are highly influential and widely respected, in the specific domain of shadow education both have produced ambiguous findings. Methodological approaches may have improved over time, but both TIMSS and PISA have suffered from difficulties in design and implementation, and some analyses have presented conclusions that are misleading.

The article draws on the work of Bray (2010), and like that paper commences with remarks about the focus of investigation. Bray included some focus on TIMSS and PISA, and the current article extends that focus. It notes the sequence of iterations in the TIMSS and PISA work, and highlights specifics of the questions relating to shadow education. The article then concentrates on ways that researchers have analyzed and presented the TIMSS and PISA data. It recognizes contributions from these studies but also shows shortcomings. The concluding section provides some indication of ways in which the shortcomings may be addressed in order to gain a clearer understanding of the scale and implications of shadow education around the world.

### Defining the Focus of Investigation

The opening paragraph of this article commenced with a term that at first sight may seem clear but on closer scrutiny may be problematic. This

<sup>3</sup> For the first assessment, four countries on what TIMSS (Gonzales and Smith 1997, chap. 1, 3) called "a Southern Hemisphere school schedule" (even though one country was just north of the equator and another was much further north) conducted the assessment in 1994. These countries were Australia, Korea, New Zealand, and Singapore. Nevertheless, for convenience the assessment is usually called TIMSS 1995 rather than TIMSS 1994/95.

term is “private supplementary tutoring,” which researchers have explicitly or implicitly defined in different ways. With explicit definitions, it is at least possible to see clearly the variations in foci of different researchers, but with implicit definitions the variations are blurred and thus even more problematic.

The first question to answer is what is meant by “private.” Many people (e.g., Bray 1999; Silova et al. 2006; Aslam and Atherton 2012) have defined “private” in a financial light, that is, whether consumers have to pay directly for the service. In this perspective, private tutoring is distinguished from tutoring that is provided free of charge by family members, community bodies, and others. However, other people, including the persons receiving tutoring, may understand the word differently. They may consider the tutoring to be private if it is received privately, that is, away from classmates and outside the public space. Such an interpretation could apply to tutoring received from family members at home and also to fee-free tutoring by community bodies. Indeed even at school, when teachers provide separate tutoring beyond the hours of normal schooling, the pupils and teachers may view it as private tutoring even if the teachers receive no extra remuneration. Accordingly, the word “private” can have at least three different meanings in relation to tutoring: (1) private as fee paying, (2) private as taking place in a private location such as the home of the tutor or tutee, and (3) private as one-to-one teaching. All these definitions have borders that may be blurred. For example, payments may be described as contributions rather than fees, and may not always be in cash. Further, private settings may be taken to include public spaces such as libraries as well as private homes, and tutoring in the privacy of a home can be provided by publicly employed tutors. Given these diverse and complex understandings of the term “private,” care is needed in all circumstances and perhaps particularly when working across cultures in cross-national studies.

The word “supplementary” also needs scrutiny. As noted above, some forms of private tutoring supplement the regular curriculum with additional content rather than just mimicked content. This may be academic, in the form of elaborated science, literature, mathematics, and so on; it may be allied to school-sponsored forms of extracurricular activities such as sports and music; or it may be totally beyond the sphere of the school, such as religious instruction or learning of minority languages. In addition, in some societies growing numbers of supplementary providers offer training in study skills, leadership, and so on (see, e.g., Oller and Glasman 2013, 81).

The third word in the phrase is “tutoring.” Most commonly, this is taken to mean one-to-one or small-group instruction. In the context of supplementary lessons, the term can include full classes and even lecture-style instruction to over 100 students. This is common in Hong Kong (e.g., Kwo and Bray 2011), where lectures may be relayed to overflow rooms by video

and where classes may be recorded for separate showing. Technologies have brought additional forms of tutoring through computer-assisted learning that may or may not include live contact with a tutor.

These terms are in turn linked to shadow education. Marimuthu et al. (1991, 5), who were among the first users of the term in the literature, equated shadow education with what in Malaysia is called private tuition, which in turn was defined as “learning activities for the clientele of the formal school which take place outside the regular school instruction program for a fee or as a community service” (1). Along the same lines, Stevenson and Baker (1992, 1639) defined shadow education as “a set of educational activities that occur outside formal schooling and are designed to enhance the student’s formal school career,” and added that they were concerned about two groups of activities. One group occurred mainly during the period of secondary schooling and included private cram schooling, correspondence courses, and individual tutorial sessions, while the other group occurred immediately after secondary schooling in institutions known as *yobiko*, which prepared students intensively to resit for university-entrance examinations. Other researchers have excluded *yobiko* on the grounds that they serve students who have left school rather than students who are still at school. Stevenson and Baker added that their conception of shadow education comprised “activities that are firmly rooted within the private sector” (1643). This is consistent with the definitions of many other authors (e.g., Aslam and Atherton 2012; Bray and Lykins 2012; Bregvadze 2012), but not all. Indeed Baker et al. (2001, 2) presented a conception that could include public sector supplementary tutoring as well as private sector work. In that paper, Baker et al. stated that researchers “refer to the host of structured outside school achievement activities as shadow education.” They added that the term “conveys the image of outside-school learning activities paralleling features of formal schooling used by students to increase their own educational opportunities,” without indicating who covered the costs of such activities (2).

### Securing Cross-National Data through TIMSS and PISA

Both TIMSS and PISA have made very significant contributions to education systems, enabling policy makers and practitioners to benchmark their work in a cross-national context. Evolutionary advances in the studies over the decades have also made major conceptual contributions to the academic field of comparative education. This section notes iterations of the pair of studies and, for the particular focus of the current article, highlights the questions that they asked about extra lessons including private tutoring. The methodological challenges will be explained by reference to TIMSS and PISA in turn. Appendix table A1 summarizes the observations.

5. **During the week, how much time before or after school do you usually spend...**

*Circle one letter, A, B, C, D, or E, for each line.*

	<i>no time</i>	<i>less than 1 hour</i>	<i>1-2 hours</i>	<i>3-5 hours</i>	<i>more than 5 hours</i>
a) taking <extra lessons/cramming school> in mathematics? .....	A	B	C	D	E
b) taking <extra lessons/cramming school> in science? .....	A	B	C	D	E

FIG. 1.—Section of question 5, international version of TIMSS 1995 student questionnaire for population 2.

*TIMSS Design and Implementation*

The Trends in International Mathematics and Science Study evolved from the Third International Mathematics and Science Study, which itself followed the first and second studies of achievements in mathematics and in science (IEA 2013). The First International Mathematics Study was conducted in 1964, and the First International Science Study in 1970/71. Over the decades, the studies expanded in geographic coverage, and the Third International Mathematics and Science Study in 1995 focused on achievements across five grades in 46 countries and independent education systems (Gonzales and Smith 1997; IEA 2013). The survey was repeated in 1999 in 39 countries and education systems under the label TIMSS-R (Third International Mathematics and Science Study—Repeat). Thereafter, the surveys retained a steady pattern in 2003, 2007, and 2011 under the label Trends in International Mathematics and Science Study (IEA 2013).

TIMSS 1995 focused on three population groups. Population 2 was defined as “students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing—seventh- and eighth-grade students in most countries” (Gonzales and Smith 1997, chap. 1, 4). Survey instruments were translated into multiple languages to fit the countries served. In addition to booklets with test items in mathematics and science, each selected student completed a student questionnaire that collected contextual information on demographics, home environment, school climate, and perceptions and learning attitudes of students. The questions about extra lessons in the international version of the TIMSS 1995 student questionnaire are reproduced in figure 1. National research teams were permitted to choose their own wording for the parts in angle brackets (i.e., < . . . >) according to common vocabulary in their jurisdictions. Thus,

even the English-language questionnaires were permitted to have slight variations around the world.

The first positive feature of the question reproduced in figure 1 is that it existed; that is, the TIMSS leadership recognized that much organized learning in academic subjects occurred outside school hours as well as within them. A second positive feature lay in the comparisons that it permitted among locations and systems, of mathematics and science, and of durations of extra lessons. In 1999, TIMSS-R retained the question unchanged. The fact that data were collected on the same instruments in two different years gave the additional benefit of comparability over time.

However, the question was in some respects problematic. First, it asked about time usually spent on extra lessons/cramming school “during the week” and did not specify what sort of week. Research in other contexts has shown considerable variation according to proximity to major examinations and with respect to term time or vacations. For example, a survey of grade 9 students in Hong Kong found that students spent an average of 5.44 hours on private supplementary tutoring during the examination season but 3.92 hours during ordinary term time and 2.52 hours during vacations (Bray 2013, 22). Second, despite the efforts of TIMSS researchers to phrase, translate, and adapt questions with care (Maxwell 1996; Ercikan and McCreith 2002; Robitaille and Beaton 2002b, 16), respondents may not all have understood the question in the way that was intended.

For TIMSS 2003, both the population and the question were slightly modified. Population 2 was defined as “the upper of the two adjacent grades with the most 13-year-olds” (Mullis et al. 2003, 6). In most countries this meant grade 8, which became the common vocabulary for TIMSS 2003 in contrast to grades 7 and 8 in TIMSS 1995. The question on mathematics for grade 8 students, which was repeated in the same format for science, is reproduced in figure 2.<sup>4</sup> The question was an advance over its predecessor insofar as it covered the whole school year rather than just an unspecified week. However, the seasonal variations would again have been problematic. Since the question was only about “this school year,” which was still in progress, presumably the respondents excluded from consideration any lessons or tutoring that had occurred during the long vacation before the beginning of the school year. Yet the responses would still have required some averaging—and would have omitted the intensity of tutoring that is more likely to have occurred toward the end-of-year examinations.

<sup>4</sup> TIMSS 2003 had two versions of the student questionnaire (see <http://timss.bc.edu/timss2003i/context.html>), for countries in which students were taking science as an integrated subject or for countries in which students were taught in separate strands for biology, earth science, chemistry, and physics. Figure 2 reproduces the question from the integrated (“general”) questionnaire. The question for the separate strands was formatted differently but was essentially the same. Some studies (e.g., Park 2013; Song et al. 2013) have analyzed data on tutoring from both versions of questionnaires, while Huang (2013), for example, only focused on countries that taught science as an integrated subject and therefore drew only on data from the general questionnaire.

**18**

A. During this school year, how often have you had extra lessons or tutoring in mathematics that is not part of your regular class?

Fill in **one** circle only

- Every or almost every day ----- ①  
 Once or twice a week ----- ②  
 Sometimes ----- ③  
 Never or almost never ----- ④

FIG. 2.—Question 18 in the international version of the TIMSS 2003 grade 8 student questionnaire

Comments below will highlight some of the problems that have arisen during the processing of data from these 1995 and 2003 questions. The problems were rooted in both the original questions and the translations, and seem to have been recognized by the TIMSS authorities. Perhaps unfortunately, the TIMSS authorities responded by dropping the question from the TIMSS international version rather than by improving it. No questions about extra lessons, cram schooling, or tutoring were asked in the international version of the TIMSS iterations in 2007 or 2011.

#### *PISA Design and Implementation*

PISA resembles TIMSS in having a focus on educational achievements, but it differs in the scope of those achievements and in the grades covered. PISA has assessed mathematics, science, and reading literacy, and focuses only on 15-year-olds. PISA assessments have been made every 3 years, in multiple education systems: 43 in 2000, 41 in 2003, 57 in 2006, 75 in 2009, and 65 in 2012.

PISA, like TIMSS, includes student questionnaires as part of its work. The questionnaires provide “interpretive frameworks” (OECD 2009, 151) in relation to students’ achievement, and collect student-level data on various aspects of teaching and learning. Questions on extra lessons were included from the outset. In 2000, students were asked in two separate questions about “special courses” attended in their schools and outside their schools. Figure 3 reproduces the question about special courses outside their schools. It asked students to think about the preceding 3 years and to indicate whether they had received courses in the test language or other subjects, extension courses, remedial courses, training, and private tutoring. Again, a positive feature was that this question existed and thus recognized that much learning occurred outside schools. However, the respondents may have had difficulty distinguishing between the different categories. For example, the item “courses in <test language>” seems to overlap with another option “<Re-

**Q 24 During the last three years, have you attended any of these special courses outside of your school to improve your results?**

(Please <tick> only one box on each row.)

	No, never	Yes, sometimes	Yes, regularly
a) courses in <test language>.....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>
b) courses in other subjects.....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>
c) <Extension> or additional courses .....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>
d) <Remedial> courses in <test language>.....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>
e) <Remedial> courses in other subjects .....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>
f) Training to improve your study skills .....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>
g) <Private tutoring> .....	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>	<input type="checkbox"/> <sub>3</sub>

FIG. 3.—Question 24 in the international version of the PISA 2000 student questionnaire

medial> courses in <test language>.” Coverage of the last 3 years avoided pinpointing just one moment in time, but respondents may have had difficulty thinking back so far and deciding between “sometimes” and “regularly,” particularly if their patterns for the same subject could have been described as sometimes in one year and regularly in another year.

In 2003, the questions were modified a little (fig. 4). The vocabulary referred to enrichment rather than extension, and classes rather than courses. Two of the six items specified that the time was spent at school, one specified that it was out of school, and the remaining three did not specify a location. Respondents were asked to estimate the number of hours per week, and to include time on the weekend. The question requested precise numbers rather than “sometimes” or “regularly,” but it did not give any guidance on seasonal variations and would have required respondents to calculate averages. The question was in the general part of the questionnaire, and thus did not ask about specific subjects in which respondents received extra lessons. Item *d* was “work with a <tutor>” but did not indicate fee paying or free, or one-to-one or small groups. In addition, some students may have included nonacademic subjects such as music and art in their responses.

The single question in 2003 was split into two questions in 2006, and retained as two questions in 2009. Although the formats of the 2006 and 2009 questions differed, for reasons of space the following commentary focuses only on the 2009 questions. In 2012 the student questionnaire again asked only a single question fully on the topic, though with two subsidiary questions in another section.

The following question asks about the time you spend studying and doing different kinds of homework outside of your regular classes. This should include all of your studying and homework.

**Q29 On average, how many hours do you spend each week on the following?**

*When answering include time at the weekend too.*

- a) Homework or other study set by your teachers ..... \_\_\_\_ hours per week
- b) <Remedial classes> at school ..... \_\_\_\_ hours per week
- c) <Enrichment classes> at school ..... \_\_\_\_ hours per week
- d) Work with a <tutor> ..... \_\_\_\_ hours per week
- e) Attending <out-of-school> classes ..... \_\_\_\_ hours per week
- f) Other study ..... \_\_\_\_ hours per week

FIG. 4.—Question 29 in the international version of the PISA 2003 student questionnaire

Figure 5 reproduces question 31 in the international version of the 2009 student questionnaire. It asked about out-of-school-time lessons in the test language, mathematics, science, and other school subjects, with separate questions for enrichment and remedial purposes. Students were asked what types of out-of-school-time lessons they attended “currently.” Again, this question would not have captured the seasonal variations, but it was perhaps more precise than the word “typically” used in 2006, and it was arguably also an advance over 2003 when no descriptors had been used. Also, respondents were told that these extra lessons could include ones received on the school premises, although outside normal school hours. However, respondents would have had to read the text carefully to grasp this instruction, and the phrasing about out-of-school time might have misled some respondents to think about out-of-school premises. In this respect, perhaps the design of the 2003 question was clearer. Third, as in 2003 respondents were asked separately about enrichment and remedial lessons, which had been dropped in 2006 (and was dropped again in 2012). Respondents were left to decide for themselves what the words meant, but they had the possibility of indicating that they took both enrichment and remedial lessons in the same subject.

Turning to question 32 (fig. 6), again respondents had to grapple conceptually with a question that asked about out-of-school-time lessons that could nevertheless be received at school. They were then asked to indicate the time “typically” spent on such lessons, which again would be difficult to answer for an activity with seasonal fluctuation. The question did permit

**Q31 What type of <out-of-school-time lessons> do you attend currently?**

*These are only lessons in subjects that you are also learning at school, that you spend learning extra time outside of normal school hours. The lessons may be given at your school, at your home or somewhere else.*

*(Please tick only one box in each row)*

	Yes	No
a) <Enrichment lessons> in <test language>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
b) <Enrichment lessons> in <mathematics>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
c) <Enrichment lessons> in <science>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
d) <Enrichment lessons> in other school subjects	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
e) <Remedial lessons> in <test language>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
f) <Remedial lessons> in <mathematics>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
g) <Remedial lessons> in <science>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
h) <Remedial lessons> in other school subjects	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>
i) Lessons to improve your <study skills>	<input type="checkbox"/> <sub>1</sub>	<input type="checkbox"/> <sub>2</sub>

FIG. 5.—Question 31 in the international version of the PISA 2009 student questionnaire

some quantification of the lessons received, this time by ticking preset boxes rather than writing a number in the format requested in 2003. As in previous iterations, neither question 31 nor question 32 asked what sort of person provided the out-of-school-time lessons—such as a teacher, university student, or employee of a tutoring center—or whether the lessons were paid or free of charge.

### Analyses in the Research Literature

Having explained the nature of TIMSS and PISA, and the questions that were asked about extra lessons and private tutoring in a number of iterations, it is pertinent to review some analyses of the data sets to identify core messages. Again this is done in sequence, commencing with TIMSS and then turning to PISA.

#### *Analyses of the TIMSS Data*

One of the first presentations on shadow education from the TIMSS data, based on the 1995 round, was by Baker et al. (2001). Their paper took the

**Q32 How many hours do you typically spend per week attending <out-of-school-time lessons> in the following subjects (at school, at home or somewhere else)?**

*These are only lessons in subjects that you are also learning at school, that you spend learning extra time outside of normal school hours. The lessons may be given at your school, at your home or somewhere else.*

*(Please tick one box in each column)*

	<test language>	<Mathematics>	<Science>	Other subjects
I do not attend <out-of-school-time lessons> in these subjects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Less than 2 hours a week	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 or more but less than 4 hours a week	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 or more but less than 6 hours a week	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 or more hours a week	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

FIG. 6.—Question 32 in the international version of the PISA 2009 student questionnaire

data set of responses to question 5a (i.e., only mathematics and not science) and investigated various dimensions. Their study included the bar chart reproduced as figure 7. The text stressed that the bar chart showed any amount of shadow education, that is, weighting equally students who responded less than 1 hour and more than 5 hours. On a minor terminological note, one might query the term “nation” to describe some entities including Hong Kong, the French education system of Belgium, and the Flemish education system of Belgium.<sup>5</sup> More substantively, the bar chart appeared to show that Colombia, Latvia, Slovak Republic, Philippines, and South Africa had considerably higher participation rates in shadow education than Japan, Hong Kong, and Korea. To persons knowledgeable about these societies, the picture seems counterintuitive to the point of deserving fundamental scrutiny. Nevertheless, the paper simply presented the data as empirical fact.

Queries of the picture presented in figure 7 might have been based on several grounds. One could have concerned sampling, since a number of

<sup>5</sup> The Latvian sample was also partial. The bar chart refers to “Latvia (LSS).” TIMSS documentation (Gonzales and Smith 1997) indicated that LSS meant Latvian speaking schools—chiefly in contrast to Russian speaking schools.

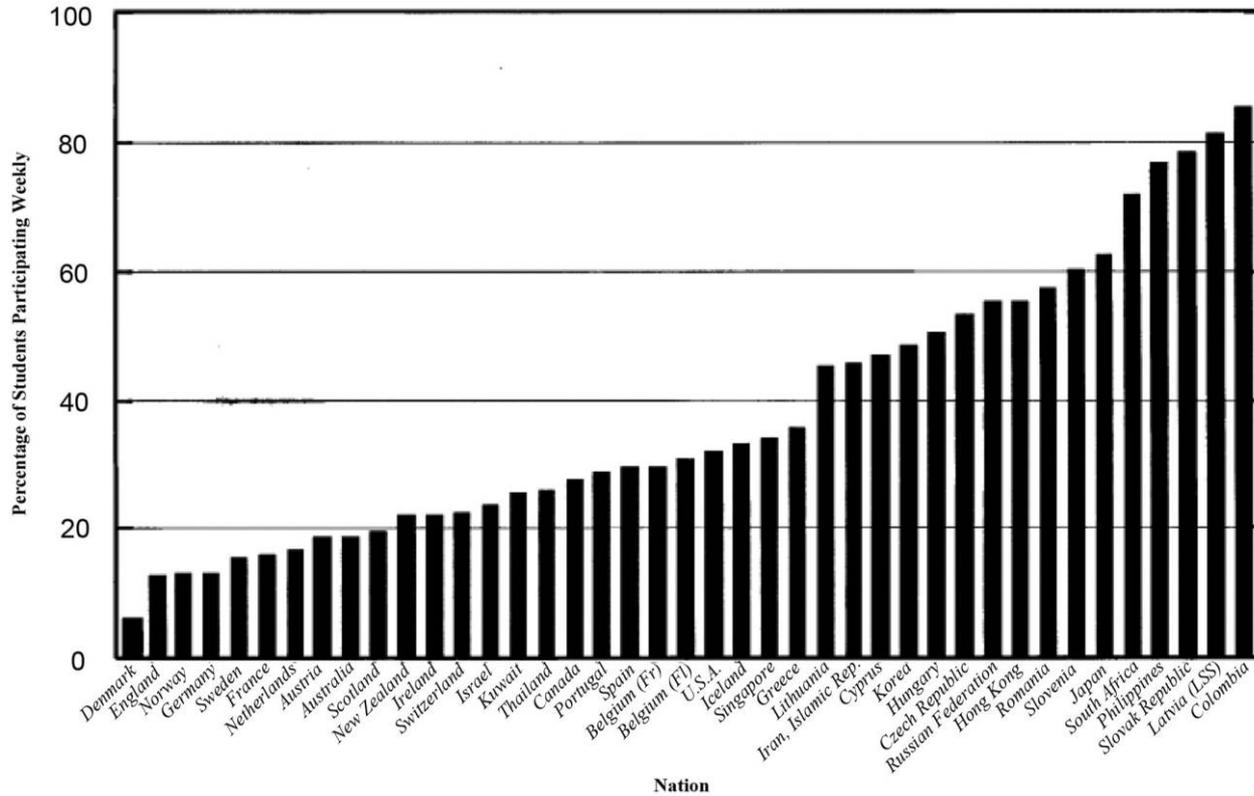


FIG. 7.—Shadow education participation in mathematics as reported by Baker et al. (2001, 7) on the basis of TIMSS 1995

country data sets did not satisfy the TIMSS guidelines.<sup>6</sup> More relevant to the current article would have been reflections on the nature of the question and on ways in which it might have been understood by the respondents. In addition to the matters already highlighted in the English version could have been problems of translation. Baker et al. (2001, 5) stated that “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.” The current authors sought the Chinese, Japanese, and Spanish versions from the IEA authorities but were unable to obtain them and so have been unable to form an independent opinion. However, the literature has indicated that at least one Spanish version was problematic. Wolf (2002, 332), when preparing a similar analysis of the data for science, acknowledged a translation error “in the item regarding ESI [extra school instruction] in Colombia” and for this reason excluded Colombia from his analysis. This information deserves particular attention since the very high proportion shown for Colombia in figure 7—far exceeding Japan and Korea, for example—was among the strikingly counterintuitive features.

Moreover, the statement in the paper that the item “was found to capture the full meaning of shadow education” might need to be confronted with a question about what shadow education meant precisely to the authors, let alone to the many people who had played roles in producing the TIMSS data. For example, the education systems of Japan and Korea during the mid-1990s were very different from those in Latvia and Slovak Republic, and the question may have measured different phenomena. In the former pair of countries, “cramming school” was presumably translated as *juku* and *hagwon*, which were clearly identifiable institutions well known to secondary students. In Latvia and Slovak Republic, parallel out-of-school institutions were uncommon, and it is more likely that respondents had in mind extra lessons that were part of the routine work of schools.

Going further, the paper by Baker et al. not only presented these data at face value, but it then made interpretations in an arbitrary way. The paper recognized that the TIMSS questionnaire “did not ask students why they used shadow education” (2001, 5). Lacking this information, the authors decided to make their own judgments. They did this by regressing on students’ mathematics scores a dichotomous indicator of students’ reported time in extra lessons and including a set of control variables. One part of their conclusions

<sup>6</sup> Gonzales and Smith (1997, chap. 3, 10) stated that data from Australia, Austria, Belgium, Bulgaria, Netherlands, and Scotland did not satisfy guidelines for sample participation of grade 8 students in the written assessment. In addition, data from four countries did not meet the age/grade specifications and had a high percentage of older students, data from six countries had unapproved sampling procedures at the classroom level, and data from one country had an unapproved sampling procedure at the school level.

that has been widely quoted concerns whether the extra lessons could be viewed as for remedial or enrichment purposes.<sup>7</sup> The paper classified the countries as dominated by extra lessons that were remedial “if substantially more low math ability students than high math ability students participate” (8). Conversely, other countries were classified as being dominated by an enrichment motivation; and a few countries were classified as mixed on the grounds that remedial and enrichment were in similar proportions.

This type of classification might seem logical from an overview perspective. However, students themselves (and their parents) do not make self-assessments against means for the whole country in which they live. Rather, their point of reference is their own classroom and perhaps their own school. Further, their classifications have an important time dimension. A student who has been at the top of the class but slips to fifth place may experience parental pressure to undertake “remedial” tutoring to regain the place at the top of the class. Thus, the definition of remedial or enrichment by the students themselves is likely to be very different from that of the external researchers looking across national and global data. Also, as reflected in the PISA questionnaires, students may in the same block of time take some extra classes for remediation and others for enrichment.

Wolf’s (2002) publication, mentioned above, deserves further attention since it paralleled the work by Baker et al. (2001) with a focus on TIMSS 1995 data in mathematics and science. Wolf explicitly avoided the term “shadow education,” considering it “somewhat ambiguous” (2002, 331). He preferred the term “extra-school instruction,” although did not in his chapter ask how “extra lessons in mathematics,” which was the phrase used in his transcription of the English version of the questionnaire (332), would have been understood by the respondents across the countries and education systems. Wolf presented a pair of tables showing reported data on extraschool instruction for the lower grade of population 2 (i.e., grade 7) in mathematics and science. The first table, for mathematics, presented data on 31 countries and systems, and the second table, for science, presented data on 32 countries and systems.<sup>8</sup> At the beginning of his chapter, Wolf indicated that “not all countries administered the items regarding extraschool instruction” (332). However, this statement was inconsistent with the IEA’s documentation.<sup>9</sup> And,

<sup>7</sup> See, e.g., Lee et al. (2009), Southgate (2009), Askew et al. (2010), and Song et al. (2013).

<sup>8</sup> Hungary was included in the science table but omitted without explanation from the mathematics table. It seems odd, since the question itself covered both subjects (see fig. 1), and it seems unlikely that the Hungarians would have deliberately omitted the part about mathematics while retaining the part about science.

<sup>9</sup> The TIMSS user guide (International Study Center 1995, v) stated that no population 2 student background data for Bulgaria were included in the international data set. The document also indicated (sec. 4, 2) concerning the French data that the section of question 5 on working at a paid job (not shown in fig. 1 of the current article) either was not administered or delivered data that were not internationally comparable. However, the document was silent about the sections of question 5 on extra lessons, implying that they were included in data collection for all countries and systems except Bulgaria.

as noted, Baker et al. had used the same data set to present statistics from 41 countries and systems.<sup>10</sup>

On the basis of the data, Wolf noted not only the wide range of patterns but also that no country showed an absence of extraschool instruction. Like Baker et al. (2001), Wolf correlated patterns of extraschool instruction with achievement scores and concluded that extraschool instruction was “used more for remediation than enrichment” (2002, 340), even though the views of the learners (or their parents) had not been sought. Wolf added that extraschool instruction was “a phenomenon that cannot be ignored,” especially because it “requires a real commitment on the part of both parents and students since it is not publicly financed” (338). This last statement was not derived from the TIMSS data. The subsequent paragraphs in Wolf’s chapter focused on Japan and Korea and certainly there were grounds to assume that much extraschool instruction in those countries was not publicly financed. However, it seems likely that a significant proportion of extra lessons were free of charge even in Japan and Korea,<sup>11</sup> and proportions are likely to have been greater still in many other countries.

Turning to TIMSS 2003, Song et al. (2013) are among researchers who have analyzed data on question 18 (fig. 2). They wrongly interpreted the data to refer to private tutoring—even including that term in the title of their paper—when the question actually asked only about “extra lessons or tutoring . . . that is not part of your regular class.” Song et al. also wrongly stated (125 and again 129) that Baker et al. (2001) had focused on private tutoring, even though Baker et al. had not in that paper included the private element in their definition of shadow education. Song et al. may have been misled by the different definition of shadow education presented by Stevenson and Baker (1992), which, as noted above, did focus on “activities that are firmly rooted within the private sector” (1643).

A major thrust of the paper by Song et al. was comparison of patterns in two pairs of countries (*a*) with “high-school-quality levels” (as indicated by mathematics scores on various TIMSS 2003 questions) and “high participation rates in private tutoring” (as indicated by responses to the TIMSS 2003 question 18 in the student questionnaire) and (*b*) with “low-school-quality levels” and “high incidences of private tutoring” on the same indi-

<sup>10</sup> However, the paper by Baker et al. focused on the whole of population 2 (i.e., grades 7 and 8). For a reason that was not explained, the pair of tables in Wolf’s chapter focused only on grade 7. Wolf indicated (2002, 333) that he had also looked at grade 8 and that results for that grade were “quite similar” to those for grade 7.

<sup>11</sup> With reference to Japan, Matsuoka (2014, 11), making reasonable assumptions on the basis of PISA 2006 data, stated that 13.6 percent of sampled grade 10 students received fee-paying extra lessons in mathematics, while about 18 percent received fee-free extra lessons in mathematics. These numbers applied to grade 10 rather than grade 7 and to 2006 rather than 1995, but it seems likely that even in Japan a significant proportion of TIMSS 1995 respondents were receiving fee-free lessons. In Georgia, 2011 TIMSS data collected by one of the authors (Kobakhidze) through a supplementary national question in the international questionnaire of grade 8 students indicated that 16.3 percent of students receiving extra classes did not pay for those classes.

cators. The countries in the former category were Korea and Taiwan, and those in the latter category were Philippines and Romania.<sup>12</sup> Again, the apparent patterns presented by the data were arguably counterintuitive. A graph presented by Song et al. (2013, 130) suggested that Ghana had the highest incidence of private tutoring [extra lessons] in mathematics, reaching nearly 90 percent of grade 8 students. Next on the list, in order, were Egypt, Romania, Armenia, Botswana, South Africa, Philippines, Morocco, and Lebanon. Only after this group of nine did Korea appear, and Japan was ranked twenty-eighth out of 45. A portrait of Japan having less private tutoring (46 percent of students) than Botswana (69 percent) does not seem consistent with other portraits of those countries (e.g., Ministry of Education and Training 2008, 13; Paviot 2010, 9).<sup>13</sup> Broader awareness of patterns could have sent a signal to Song et al. that the data might contain public tutoring as well as private tutoring, although even with allowance for this inclusion the figure for Botswana looks too high. Similar doubts might be raised on whether Romania and Philippines were really the third highest and seventh highest among the 45 countries.<sup>14</sup>

#### *Analyses of the PISA Data*

While TIMSS dropped the questions about tutoring/extra lessons after 2003, PISA maintained questions on the topic. Both the OECD itself and the academic community have undertaken various analyses of the data, among which a selection may illustrate the benefits and challenges.

Beginning with OECD work, a 2011 report drew on PISA 2006. It commenced (OECD 2011, 22–23) with a cautionary statement that could usefully be read by other authors cited in this article:

<sup>12</sup> The description of Taiwan as a country would not be universally accepted. It is used in this sentence because it was the word used by Song et al.

<sup>13</sup> The Japanese study stated that *juku* served 65.2 percent of grade 9 students and that an additional 6.8 percent of grade 9 students received tutoring at home and 15.0 percent followed correspondence courses. These figures were for all subjects rather than specifically for mathematics. Nevertheless, they are strikingly higher than the figure for Japan reported in the TIMSS 2003 mathematics grade 8 data (which included fee-free extra lessons). Paviot was reporting on a 15-country study conducted under the auspices of the Southern and Eastern Africa Consortium for Monitoring Educational Quality. It focused on grade 6 rather than grade 8 and on reading as well as mathematics. Nevertheless, Paviot reported that only 5.9 percent of surveyed pupils in Botswana were receiving paid [supplementary] lessons. This was at striking variance with the 69 percent figure presented by Song et al., even though the actual PISA question from which the figure was derived included fee-free as well as fee-paying extra lessons.

<sup>14</sup> Concerning Romania, Song et al. (2013, 129) cited a 2000 UNESCO report that itself cited a 1994 study of grade 12 students in a national sample, among whom 32 percent in rural areas and 58 percent in urban areas had received private supplementary tutoring. More recent studies have shown lower figures. For example, the stratified random survey of secondary students by Metro Media Transilvania and Agentia pentru Strategii Guvernamentale (2007) found that 27 percent received tutoring in Romania. Concerning Philippines, Song et al. (2013, 129) merely stated that “a number of Filipino students” were known to enroll in review centers or to hire private tutors. Other parts of the paper cited de Castro and de Guzman (2010, 2012), whose work does show that significant amounts of private tutoring existed in parts of Philippines but does not seem to support the 74 percent figure reported from TIMSS 2003, even after allowance for other forms of (public) extra lessons.

The nature of out-of-school lessons is not necessarily the same across countries and even within countries. Even when students report attending the same type of out-of-school-time lessons, they might attend a lesson with different incentives and objectives. For example, some schools offer additional lessons to students who need remedial education, while other schools offer them to students who seek further enrichment. In some countries, out-of-school-time lessons with school teachers are systematic and standardised across schools, while in other countries these are organised by individual schools and the quality of lessons varies greatly from school to school.

As practice varies among and within countries, it is difficult to generalise about the differences between out-of-school-time lessons with school teachers and those with non-school teachers. In some countries, tuition fees are charged for out-of-school-time lessons with non-school teachers, but fees are not charged for lessons with school teachers. In other countries, the opposite is true, or the setup of out-of-school time lessons with any kind of teacher may be entirely different. . . . Given that out-of-school-time lessons across and within countries vary so much, and because this variability is not captured in the data in this report, it is impossible to generalise about the effects of out-of-school-time lessons.

This cautious approach indeed seems justified and more valuable than overconfident assertions. The report provided 24 charts and 70 annex tables to show patterns and variations. The conclusion pointed out that “simply adding hours to the school day or encouraging students to spend more time in after-school lessons or individual study would not automatically help low-performing countries to improve their test scores” (79). Instead, education authorities were recommended to find ways to improve the quality of lessons. The report also stressed the need for tighter research focus on the theme.

Another OECD report has focused on Serbia and has drawn on PISA 2006 and PISA 2009 data (OECD 2012). Some data were cross-national, to show Serbia in comparison with other countries, and other data focused only on Serbia. Figure 8 exemplifies the latter, showing data derived from question 31 of the PISA 2009 student questionnaire. It shows contrasts according to the respondents’ reported views on whether their lessons were for enrichment or remediation, the different subjects, and the streams in which students were enrolled (general, vocational, and vocational directly to labor market). However, careful comparison of this table with the wording of question 31 (fig. 5) again shows inaccuracies.

- The title of the table refers to out-of-school lessons, but question 31 was about out-of-school-*time* lessons, which “may be given at your school, at your home or somewhere else.”
- Question 31 did not ask about enrichment lessons in other subjects. It did ask about “lessons to improve your <study skills>,” but that is not the same.

Elsewhere, too, this 2012 report was less careful than its predecessor (OECD

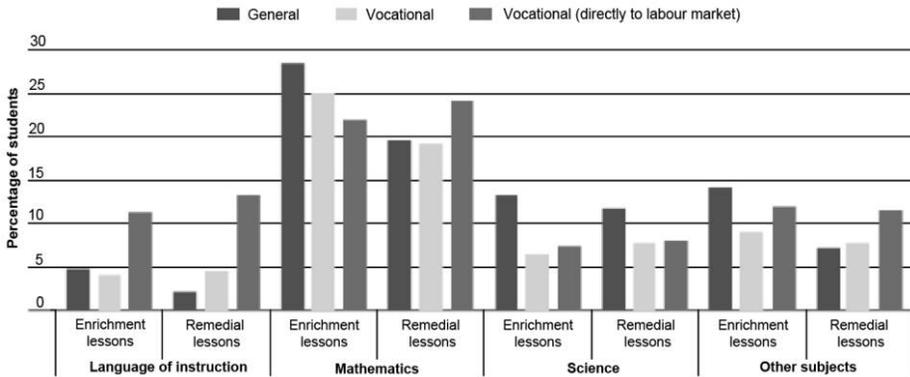


FIG. 8.—Share of students attending out-of-school lessons, by subject and type of lesson, Serbia. NOTE.—This title is copied precisely from the figure title in its source (OECD 2012, 48), except that “Serbia” has been added. However, the title is not an accurate reflection of the question from which the data were derived.

2011). It referred variously to private tutoring, out-of-school lessons, and shadow education in ways that were not always clearly defined. It did define “remedial” and “enrichment,”<sup>15</sup> but it did not note that those definitions had not been presented to the respondents when they were answering the questions. Nor did the report comment on the challenges that respondents might have faced in answering the questions even if the definitions had been presented to them. The authors presented policy recommendations for the Serbian government on various matters including shadow education. Yet as the report noted, “PISA does not provide information about where [out-of-school] lessons take place, how students are selected for out-of-school lessons, whether they happen regularly and whether students pay for these lessons” (OECD 2012, 48).

Comparable shortcomings may be found in various academic works. For example, Southgate (2009, 1) commenced by defining shadow education as “the use of privately funded assistance in school subjects” and by implying that this definition was derived from both Stevenson and Baker (1992) and Baker et al. (2001) even though the latter, as noted, had a definition that could permit publicly funded as well as privately funded supplementary lessons. Southgate then proceeded to analyze data derived from sections *d* and *e* of question 29 in PISA 2003 (fig. 4) as if they did indeed exclusively show privately funded activity. Lee et al. (2009, 904) presented a bar chart similar to that presented by Baker et al., ranking 57 countries and jurisdictions on

<sup>15</sup> “Remedial lessons are any additional lessons designed to help students with learning difficulties (‘reinforcement’), whereas enrichment lessons are designed to extend the abilities of more capable students” (OECD 2012, 45).

“the proportion of 15-year-olds who participated in private tutoring, using PISA 2006 data,” omitting recognition that the data could apply to public as well as private tutoring. Choi et al. (2012) presented a paper on “private tutoring and academic achievement in Korea” on the basis of PISA 2006 data with the same lack of recognition. And Runte-Geidel (2013) made the same error with respect to data on Spain, compounding the error (273) by placing on a single graph data from PISA 2000, 2003, 2006, and 2009 under the label “Use of Shadow Education” as if the data from the very different types of questions in each iteration could somehow be standardized on a single measure.

### Further Challenges of Adaptation, Translation, and Student Engagement

Like all cross-national surveys, TIMSS and PISA have encountered challenges of adaptation and translation of questions. Some of these difficulties have been mentioned above with reference to TIMSS, and they may be elaborated on with reference to PISA. Also noteworthy are the attitudes of students toward the processes and content of the questionnaires that have affected the reliability of data.

Adaptations to questions have been necessary even in English to allow for different vocabularies in different societies. Variations for items 31*b* and 31*f* in the PISA 2009 student questionnaire (fig. 5) include these.<sup>16</sup>

- In the United States, the questions were about “Enrichment lessons in mathematics” and “Remedial lessons in mathematics,” exactly like the international version.
- In New Zealand, the questions were about “Extension lessons in mathematics” and “Remedial lessons in mathematics.”
- In Ireland, the questions were about “Enrichment lessons in mathematics” and “Learning support in mathematics.”
- In England, the questions were about “Additional Mathematics lessons which go beyond what you have learned in your Mathematics class” and “Catch-up lessons to help you with Mathematics.”

The fact that country-level personnel felt the need to make adaptations even in the English version is a signal that the basic concept might not have been universally clear. Further difficulties arose in translation, which may be illustrated by the Chinese versions in Shanghai, Taiwan, and Macao.

- For “enrichment,” Shanghai used a pair of phrases that literally mean

<sup>16</sup> Because countries were permitted to add their own questions, the item that was question 31 in the international version was not always question 31 in the national versions. However, at least in these four countries the sub-questions were retained in the same order.

“strengthening class or interest class” (*qianghuaban huo xingquban*, 强化班 或 兴趣班). The concept of an interest class might have distracted students into thinking about nonschool subjects in which they had interest, even though the stem of the question had stressed that it was only about subjects that respondents were also learning at school. Taiwan used “strengthening course” (*jiaqiang kecheng*, 加强 课程). Low-achieving students of course need to be strengthened as well as high-achieving ones. The fact that this question was aimed at high-achieving students who were gaining further strengthening might have become clear when the respondents reached the question about remedial courses (see below), but only the diligent respondents would have gone back to change their answers if the contrast with remedial required such a change. The emphasis on classes in Shanghai and courses in Taiwan would have steered respondents away from considering one-to-one tutoring. Most problematic was the vocabulary used in Macao. The questionnaire used the term “add profit [or lubricant]” (*zengrun*, 增潤), which is not in daily usage, especially in educational matters.

- For “remedial,” Shanghai used the term *buxiban* (补习班), which literally means simply “tutoring class” and is neutral rather than implying a need for remediation. Taiwan used the term “rescue course” (*bujiu kecheng*, 補救课程), which is more evocative of the intended meaning. Macao used the term “guidance class” (*fudaoban*, 輔導班), which indeed is commonly a term for remedial class but might orient respondents more toward such classes in school rather than to out-of-school tutoring.

The fact that the Macao version used terminology for enrichment that seemed odd to the educators and ordinary people in those societies consulted by the current authors underlines the challenges in such work. If the respondents did not clearly understand the question, or did not understand it in the way that was intended, then the responses cannot be taken at face value.<sup>17</sup>

It is also important to note caution expressed elsewhere on the seriousness with which respondents may have addressed such questionnaires. Experiences in Ireland with reference to PISA 2009 illustrate the challenges. LaRoche and Cartwright (2010) investigated whether elements in an apparent decline in Ireland’s PISA scores reflected a decline in proficiency or an increase in student disengagement. By the latter was meant “the reluctance of students to participate, [and] of some students leaving the test room before the end of the test or being impatient to finish and leave” (5). Cosgrove

<sup>17</sup> Zhou and Wang (2013) pointed out that in Macao the responses to questions 31 and 32 “don’t line up,” with the number of students indicating in response to question 32 that they received out-of-school-time lessons for varying durations being much greater than the number of students who indicated in response to question 31 that they received enrichment/remedial lessons in the various subjects. Zhou and Wang remarked that the “unusual wording” had led to “confusion in answers” (11).

(2011) further analyzed the responses in the core questionnaires about reading, mathematics, and science to identify patterns of nonresponse and non-completion of questionnaires. Patterns were varied, but in mathematics, for example, Cosgrove concluded (41) that the lower scores reflected both a decline in proficiency and an increase in disengagement.

Complementing Cosgrove's quantitative analysis are interview data presented by Mac Ruairc (2011). He focused on the volume and intensity of the testing process, the strategies used by the respondents to complete the test items, and the attitudes toward personal questions.<sup>18</sup> Extracts from the interview transcripts (148–49) included:

I was trying to do me best at the start but by the time I got to page 40 odd I just thought this is never going to end. Then I kinda stopped and I didn't think as much about the answers or just picked one. It depended like.

I stuck with it for a while and then when I looked through the pages at what was left. I just thought 'oh crap' and started to tick everything.

I read some of them and tried to think logically what they were looking for. And sometimes I just ticked a few boxes if I couldn't get it.

The student background questionnaire had been given to the respondents after they had completed the content questionnaires on reading, mathematics, and science. It had 52 pages and 79 questions with many sub-questions. Adding to the problem of length was concern about the lack of anonymity and the personal nature of some questions. The lack of anonymity arose because students had been asked to write their names, and they were not convinced by the promise that the responses would be confidential.<sup>19</sup> Again, quotations from the interview transcripts (Mac Ruairc 2011, 152) illustrated the students' perspectives:

I thought it was just about us but it was about our Mas and Das [mothers and fathers] and all that. It was a bit personal, a bit nosey like. Does your Ma have a PhD? And yer [your] name was on it, and we were told it would be confidential.

I was bothered by the question about who lives at home with you. It made me feel uncomfortable. It's our business.

I think that they are saying that if you have all them DVDs and expensive stuff and not books and the schools things that your Ma doesn't really care about school or what yeh [you] do in school. There spending money on things that don't help you at school, but that's what they want to do.

Mac Ruairc did not make any specific remarks about the questions on extra lessons, which appeared on pages 21 and 22 of the 52-page booklet,

<sup>18</sup> It must be recognized that the interviewees were in a disadvantaged urban school and may have had different reactions from other students. Nevertheless, their remarks have considerable pertinence.

<sup>19</sup> The reason why they were asked to write their names was not indicated. It may have been because PISA had a separate questionnaire for parents and needed the students' and parents' questionnaires to be matched, although normally that is done through numbers rather than names.

that is, about halfway through. This location may have been before the threshold of irritation and fatigue for many students, but it was after the questions about parents' education and home possessions. Moreover, the students had already had to handle the booklets on reading, science, and mathematics, on which they had had negative perceptions. As such, these students may not have been inclined to think carefully about what was meant by "lessons outside of school time . . . in subjects that you are also learning at school" and about the meanings of "enrichment" and "remedial."

The students interviewed by Mac Ruairc, it must be stressed, were only in one school and in one country. Nevertheless, disengagement with the questionnaires was also a problem in other countries (Cosgrove 2011). The design and implementation challenges are to some extent generic and are certainly not limited to PISA. Statisticians do have ways to assess plausible responses (Rutkowski et al. 2010), but the fact remains that at least some numbers in the PISA data sets should not necessarily be taken at face value.

### Conclusion

Both TIMSS and PISA have secured very strong visibility and credibility among academics, policy makers, and members of the general public across the globe. In part, this reputation is based on the technical competence of the administering organizations and the professionals with whom they work. The reputation is also based on the authority gained from numerical data sets. Concerning TIMSS 1995, Robitaille and Beaton (2002a, 410) observed that it "was, in many respects, a highly innovative study" and expressed the hope that those who undertake studies in the future would benefit from the advances made. This has certainly been the case in many dimensions (Martin and Mullis 2012). PISA has arguably gained even stronger technical power and status (Lockheed 2013), and many would agree with Bolívar (2011, 64) that "PISA has managed to occupy the public education space like no other type of report or survey. This media success is due, without doubt, to the design of the instrument itself: worldwide application, attention paid to the validity of the instrument, periodic regularity of the survey, generation of its own data, participation that depends on the respective public authorities, policy based on evidence, [and] tests focused on competencies and not on curricular contents."<sup>20</sup> TIMSS and PISA have had a discernible impact on education systems around the world, and in the academic domain they have contributed to a fundamental shift in part of the field of comparative education (Pereyra et al. 2011; Meyer and Benavot 2013; Leung and Park 2014).

However, in some ways the credibility of TIMSS and PISA has brought dangers. This article, building on the work of Bray (2010), has shown that

<sup>20</sup> Bolívar was writing specifically about the Ibero-American setting, but these remarks also apply more widely.

in the specific domain of shadow education the data emanating from both enterprises have been problematic. Meyer and Benavot (2013, 21), making an observation about PISA that could equally apply to TIMSS, have pointed out: “The fact that this apparatus relies on numbers and statistics does not mean that it is anchored in transparent, objective, uncontestable truth. In fact the ‘cloud of data’ . . . may easily operate like a Rorschach in which anyone can find support for any preconceived idea. It creates the opposite of transparency because key assumptions and key decisions about categorization and the construction of measures are *black-boxed* by a complex array of behind-the-scene judgments and decisions.” In the specific domain of shadow education, Southgate (2009) was misled by the reputation of PISA when she wrote that PISA was “ideal” for analysis of private tutoring (35) and that it “provides a precise measure of . . . tutoring and outside-school classes” (40). This article has shown that the questions in both TIMSS and PISA have suffered from ambiguities compounded by translation problems. Also, the validity of some responses has been threatened by the negative attitudes of at least some respondents to the administrative processes and to parts of the content. Serious problems in interpretation may then arise when the data sets appear to provide precise measures but in fact do not. Misinterpretations are extended when authors report what they would have liked the questions to have asked rather than what the questions actually did ask. This article has also highlighted examples in which researchers have misrepresented the work of other researchers.

Shadow education is a phenomenon of great significance since it consumes huge expenditures and a great deal of time of both students and their families. It also has a backwash on regular schooling, may be a major determinant of student achievement, and can have far-reaching implications for social equity. Moreover, the TIMSS and PISA data have helped to show that shadow education is a global phenomenon. However, for analytical purposes, an initial problem has arisen with definitions and concepts. Some authors have defined shadow education as encompassing a wide range of organized forms of out-of-school activities, while others have restricted the focus to academic subjects that are also covered in the students’ schooling. Further divergence has arisen on whether shadow education embraces activities financed from any source, including the public sector, or whether it is restricted to activities for which the clients or their families have had to pay a fee. Moreover, the researchers’ definitions have not always matched the vocabularies and conceptions of the recipients of shadow education. In particular, the term “private tutoring” can embrace or exclude multiple forms of “private” and multiple forms of “tutoring.” Such ambiguities and divergences must be recognized by researchers when reviewing data collected by different bodies at different points in time and in different locations.

Looking ahead, a pertinent question is what might be recommended for

TIMSS and PISA. First, again they are both to be commended for having included questions on out-of-school-time lessons. Evidently the TIMSS authorities recognized the problems with the questions in the 1995 and 2003 surveys and then dropped the question altogether. It is regrettable that the question was dropped rather than improved, since indeed out-of-school-time lessons are a major occupation for students around the world, and the phenomenon appears to be growing rather than diminishing (Bray 2009, 2011; Mori and Baker 2010; Bray and Lykins 2012). In the absence of TIMSS questions in the 2007 and 2011 iterations, it is valuable to have the PISA questions in the 2006, 2009, and 2012 iterations.

The question then concerns the desirable focus of future inquiries. Researchers and policy makers who are interested only in matters of curriculum coverage and student achievement may not be strongly concerned with the identities of the persons providing the extra lessons—such as regular school teachers, amateur tutors working informally, or professional tutors employed by companies—or whether the extra lessons are provided free of charge or for a fee. However, these matters are of major interest to researchers and policy makers concerned with efficiency in education systems and equity in wider society, and cross-national data of the sort that can be provided by TIMSS and PISA could indeed be very valuable. The OECD (2011, 23) has rightly indicated that other important factors related to out-of-school-time lessons would include the quality of teaching, the resources used during lessons (textbooks, school materials, etc.), and the motivations of students to participate in the lessons. If clear and well-targeted questions can be developed and then held constant over several iterations, then TIMSS and PISA would also provide valuable longitudinal data.

On one more specific aspect, perhaps the efforts to distinguish between extra lessons for enrichment and for remediation are less useful. This article has exposed fundamental doubts on whether the questions are meaningful when posed in this form. First, the definitions of enrichment and remedial by researchers looking at whole systems may be different from those of the learners and their families. Second, the definitions may impose an unrealistically clear-cut categorization on the content of lessons at any one time and in any one season. The OECD's 2011 report might have disappointed some readers by avoiding sharp categorizations of this sort, but nuanced understandings of the type sought by the OECD report are in reality likely to be less misleading and more useful.

A further need highlighted in this article has been for researchers to inform their interpretations with wider understanding of the education systems on which they are reporting. Some studies have simply taken the TIMSS and PISA data at face value and have arrayed the data on bar charts, tables, and graphs without having obtained sufficient background information on the countries and education systems represented. Fortunately, increasing

numbers of quantitative and qualitative studies on shadow education in individual countries are becoming available. These can be used as benchmarks against which to compare the data from TIMSS and PISA studies and to detect patterns that are counterintuitive and in need of careful checking.

Many of the challenges in cross-national studies are inherent in the nature of the undertaking, which requires negotiation and compromise between multiple actors. On the technical front of working across languages, Robitaille and Beaton (2002a, 414) observed that “there is no way to be absolutely sure that all translations had exactly the same meaning and same level of language difficulty.” Certainly the efforts undertaken by TIMSS and PISA to strengthen the quality of the work are to be applauded, but the limitations must still be recognized. Moreover, this requires informing the consumers of research as well as the technical personnel concerned with the mechanics. Tables of numbers may appear to have precision, but may have been constructed on a shaky basis.

Ending on an optimistic note, the PISA authorities decided in 2013 that the 2015 iteration would bring a radical change in format rather than mere tinkering with earlier questions and that this radical change would permit much improvement in the questions about shadow education.<sup>21</sup> Moreover, some TIMSS national research teams at the country level have recognized the need for better data in their own contexts and have added questions on shadow education in their specific settings.<sup>22</sup> In parallel, the number of independent national and subnational studies on shadow education is likely to continue to expand significantly, thereby providing complementary data that can be juxtaposed with cross-national data to form a much clearer picture of the nature and implications of shadow education.

<sup>21</sup> Information provided to Mark Bray by OECD personnel, April 19, 2013.

<sup>22</sup> For example, the 2011 iteration of TIMSS was combined with the IEA’s Progress in International Reading Literacy Study at grade 4. The Georgian and Hong Kong teams added questions on this topic.

TABLE A1  
MEASURES IN TIMSS AND PISA QUESTIONNAIRES

Assessment	Item in Student Background Questionnaire	Measure of Tutoring	Subject of Tutoring	Implied Definition of Tutoring
TIMSS 1995	Q.5: "During the week, how much time before or after school do you usually spend . . ."	Time (ranging from "no time" to "more than 5 hours")	Mathematics and science: (a) taking <extra lessons/cramming school> in mathematics, (b) taking <extra lessons/cramming school> in science	Includes all types of before- and after-school tutoring in mathematics (item <i>a</i> ) and science (item <i>b</i> ) from any provider. Asks about tutoring during an unspecified week, and thus failed to capture seasonal variations. This type of tutoring can be provided in school or outside school or both. No distinction can be made between free and paid tutoring.
TIMSS 2003	Q.18: "During this school year, how often have you had extra lessons or tutoring in mathematics [science] that is not part of your regular class?"	Regularity (ranging from "every or almost every day" to "never or almost never")	Mathematics and science: "extra lessons or tutoring in mathematics [science] that is not part of your regular class"	The question is slightly more focused on time/duration of tutoring than TIMSS 1995. Instead of an unspecified week, it is more specific about "during [the] school year." Measurement of tutoring changed from time to regularity. This understanding of tutoring excludes tutoring before the school year, that is, during long holidays. It can still include all types of tutoring with all types of providers. It can be given both inside and outside school premises. It can include free and fee-paying tutoring.
PISA 2000	Q.24: "During the last three years, have you attended any of these special courses <i>outside of your school</i> to improve your results?"	Type of courses (e.g., extension and remedial) and regularity (never, sometimes, regularly)	(a) courses in <test language>, (b) courses in other subjects, (c) <extension> or additional courses, (d) <remedial> courses in <test language>, (e) <remedial> courses in other subjects, (f) training to improve your study skills, (g) <private tutoring>	The question is about cumulative experience of tutoring during the last 3 years that might have happened outside school. Students who had less than 1 year of experience are grouped with students with 2 or 3 years of experience. Items <i>a</i> and <i>d</i> overlap. The question only asked about tutoring outside school and therefore excluded tutoring provided by the school. It does not distinguish fee-paying from free tutoring or give any information on providers.

PISA 2003	Q.29: "On average, how many hours do you spend <i>each week</i> on the following?"	Number of hours (asking for exact number of hours)	Subjects not specified: (a) homework or other study set by your teachers, (b) <remedial classes at school>, (c) <enrichment classes at school>, (d) work with a <tutor>, (e) attending <out-of-school> classes, (f) other study	On the basis of the question and instruction, the tutoring can be provided by the school or by other providers. It includes time spent on study and homework during weekdays and weekends. "Each week" refers to an unspecified week and fails to indicate seasonal variations. Averaging hours may have caused difficulty to students. Item <i>f</i> can include self-study. This question can also include tutoring in nonacademic subjects such as music and arts that are part of an official curriculum. Items <i>d</i> , <i>e</i> , and <i>f</i> can include nonschool subjects such as an additional foreign language and religious education not taught in school.
PISA 2009	Q.31: "What type of <out-of-school-time lessons> do you attend currently?"	Type of classes (e.g., enrichment or remedial)	Test language, mathematics, science, other school subjects	Question asked about extra tutoring outside (standard) school time but not necessarily outside school. It can include school-provided tutoring. The phrase "out-of-school-time lessons" might have misled some students to think about out-of-school premises. Unlike PISA 2003, PISA 2009 confined understanding of tutoring to subjects that the students were also learning at school. This could still include nonacademic subjects, such as music, art, and religion. The question introduced the notions of "remedial classes" and "enrichment classes," which were not defined for the respondents and which created difficulty in some translations.
PISA 2009	Q.32: "How many hours do you typically spend per week attending <out-of-school-time lessons> in the following subjects (at school, at home or somewhere else)?"	Number of hours (ranging from "I do not attend <out-of-school-time lessons> in these subjects" to "6 or more hours a week")	<Test language>, <mathematics>, <science>, other subjects	The question was built on broad understanding of tutoring as extra learning of academic subjects through lessons that take place outside school time. Out-of-school-time lessons can take a variety of forms; be given at school, at home, or somewhere else; and be provided free or for a fee. Again, this did not capture seasonal fluctuations.

Two additional summary tables are available online.

## References

- Askew, M., J. Hodgen, S. Hossain, and N. Bretscher. 2010. *Values and Variables: Mathematics Education in High-Performing Countries*. London: Nuffield.
- Aslam, M., and P. Atherton. 2012. "The 'Shadow' Education Sector in India and Pakistan: The Determinants, Benefits and Equity Effects of Private Tutoring." ESP Working Paper no. 38, Education Support Programme, Open Society Foundations.
- Aurini, J., S. Davies, and J. Dierkes, eds. 2013. *Out of the Shadows: The Global Intensification of Supplementary Education*. Bingley: Emerald.
- Baker, D. P., M. Akiba, G. K. LeTendre, and A. W. Wiseman. 2001. "Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement." *Educational Evaluation and Policy Analysis* 23 (1): 1–17.
- Bolívar, A. 2011. "The Dissatisfaction of the Losers: PISA Public Discourse in Ibero-American Countries." In *PISA under Examination: Changing Knowledge, Changing Tests, and Changing Schools*, ed. M. A. Pereyra, H. G. Kotthoff, and R. Cowen. Rotterdam: Sense.
- Bray, M. 1999. *The Shadow Education System: Private Tutoring and Its Implications for Planners*. Paris: UNESCO International Institute for Educational Planning.
- Bray, M. 2009. *Confronting the Shadow Education System: What Government Policies for What Private Tutoring?* Paris: UNESCO International Institute for Educational Planning.
- Bray, M. 2010. "Researching Shadow Education: Challenges and Directions." *Asia Pacific Education Review* 11 (1): 3–13.
- Bray, M. 2011. *The Challenge of Shadow Education: Private Tutoring and Its Implications for Policy Makers in the European Union*. Brussels: European Commission.
- Bray, M. 2013. "Benefits and Tensions of Shadow Education: Comparative Perspectives on the Roles and Impact of Private Supplementary Tutoring in the Lives of Hong Kong Students." *Journal of International and Comparative Education* 2 (1): 18–30.
- Bray, M., and C. Lykins. 2012. *Shadow Education: Private Supplementary Tutoring and Its Implications for Policy Makers in Asia*. Manila: Asian Development Bank.
- Bray, M., A. E. Mazawi, and R. G. Sultana, eds. 2013. *Private Tutoring across the Mediterranean: Power Dynamics and Implications for Learning and Equity*. Rotterdam: Sense.
- Bregvadze, T. 2012. "Analysing the Shadows: Private Tutoring as a Descriptor of the Education System in Georgia." *International Education Studies* 5 (6): 80–89.
- Buchmann, C. 2002. "Getting Ahead in Kenya: Social Capital, Shadow Education, and Achievement." In *Schooling and Social Capital in Diverse Cultures*, ed. B. Fuller and E. Hannum. Amsterdam: JAI Press.
- Choi, A., J. Calero, and J. Escardíbul. 2012. "Private Tutoring and Academic Achievement in Korea: An Approach through PISA-2006." *KEDI Journal of Educational Policy* 9 (2): 299–322.
- Cosgrove, J. 2011. "Does Student Engagement Explain Performance on PISA? Comparisons of Response Patterns on the PISA Tests over Time." Educational Re-

- search Centre, St. Patrick's College. [http://www.erc.ie/documents/engagement\\_and\\_performance\\_over\\_time.pdf](http://www.erc.ie/documents/engagement_and_performance_over_time.pdf).
- de Castro, B., and A. B. de Guzman. 2010. "Push and Pull Factors Affecting Filipino Students' Shadow Education (SE) Participation." *KEDI Journal of Educational Policy* 7 (1): 43–66.
- de Castro, B., and A. B. de Guzman. 2012. "From Scratch to Notch: Understanding Private Tutoring Metamorphosis in the Philippines from the Perspectives of Cram School and Formal School Administrators." *Education and Urban Society* 2 (1): 1–25.
- Ercikan, K., and T. McCreith. 2002. "Effects of Adaptations on Comparability of Test Items and Test Scores." In *Secondary Analysis of the TIMSS Data*, ed. D. F. Robitaille and A. E. Beaton. Dordrecht: Kluwer.
- Gonzales, E. J., and T. A. Smith. 1997. *User Guide for the TIMSS International Database: Primary and Middle School Years*. Chestnut Hill, MA: TIMSS International Study Center, Boston College. [http://timss.bc.edu/timss1995i/database/UG\\_1and2.pdf](http://timss.bc.edu/timss1995i/database/UG_1and2.pdf).
- Huang, M. H. 2013. "After-School Tutoring and the Distribution of Student Performance." *Comparative Education Review* 57 (4): 689–710.
- IEA (International Association for the Evaluation of Educational Achievement). 2013. "Completed Studies." IEA, Amsterdam. [http://www.iea.nl/completed\\_studies.html](http://www.iea.nl/completed_studies.html).
- International Study Center. 1995. *User Guide for the TIMSS International Database*, suppl. 3, *Documentation of National Adaptations of International Background Questionnaire Items—Populations 1 and 2*. Chestnut Hill: International Study Center, Lynch School of Education, Boston College. [http://timssandpirls.bc.edu/timss1995i/database/UG1\\_Sup3.pdf](http://timssandpirls.bc.edu/timss1995i/database/UG1_Sup3.pdf).
- Kwo, O., and M. Bray. 2011. "Facing the Shadow Education System in Hong Kong." *IIAS Newsletter*, no. 56:20.
- Kwok, P. 2004. "Examination-Oriented Knowledge and Value Transformation in East Asian Cram Schools." *Asia Pacific Education Review* 5 (1): 64–75.
- LaRoche, S., and F. Cartwright. 2010. "Independent Review of the 2009 PISA Results for Ireland." Report for the Educational Research Centre, Dublin. [http://www.erc.ie/documents/statscan\\_pisa00to09\\_final\\_report.pdf](http://www.erc.ie/documents/statscan_pisa00to09_final_report.pdf).
- Lee, C. J., H. J. Park, and H. Lee. 2009. "Shadow Education System." In *Handbook of Education Policy Research*, ed. G. Sykes, B. Schneider, and D. N. Plank. New York: Routledge.
- Leung, F., and K. Park. 2014. "Comparing Educational Achievements." In *Comparative Education Research: Approaches and Methods*, ed. M. Bray, B. Adamson, and M. Mason. 2nd ed. Hong Kong: Comparative Education Research Centre, University of Hong Kong; Dordrecht: Springer.
- Lockheed, M. 2013. "Causes and Consequences of International Assessments in Developing Countries." In *PISA, Power and Policy: The Emergence of Global Educational Governance*, ed. H. D. Meyer and A. Benavot. Oxford: Symposium.
- Mac Ruairc, G. 2011. "The PISA Girls and Ticking the Boxes: An Examination of Students' Perspectives on PISA Testing." In *PISA under Examination: Changing Knowledge, Changing Tests, and Changing Schools*, ed. M. A. Pereyra, H. G. Kotthoff, and R. Cowen. Rotterdam: Sense.
- Marimuthu, T., J. S. Singh, K. Ahmad, H. K. Lim, H. Mukherjee, S. Osman, T.

- Chelliah, J. R. Sharma, N. M. Salleh, L. Yong, T. L. Lim, S. Sukumaran, L. K. Thong, and W. Jamaluddin. 1991. *Extra-school Instruction, Social Equity and Educational Quality*. [In Malaysian.] Singapore: International Development Research Centre.
- Martin, M. O., and I. V. S. Mullis, eds. 2012. *Methods and Procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- Matsuoka, R. 2014. "School Socioeconomic Compositional Effect on Shadow Education Participation: Evidence from Japan." *British Journal of Sociology of Education*, forthcoming. doi:10.1080/01425692.2013.820125.
- Maxwell, B. 1996. "Translation and Cultural Adaptation of the Survey Instruments." In *Third International Mathematics and Science Study (TIMSS) Technical Report*, vol. 1, *Design and Development*, ed. M. O. Martin and D. L. Kelly. Chestnut Hill, MA: TIMSS International Study Center, Boston College.
- Metro Media Transilvania and Agentia pentru Strategii Guvernamentale. 2007. *Calitatea educat̄iei idin î nvăț ământul preuniversitar: Studie realizat̄ în rândul populaț iei de elevi din î nvăț ământul preuniversitar (clasele VII–XII)* [The quality of pre-university education: A study of the pupil population of preuniversity (grades VII–XII)]. Bucharest.
- Meyer, H. D., and A. Benavot. 2013. "PISA and the Globalization of Education Governance: Some Puzzles and Problems." In *PISA, Power and Policy: The Emergence of Global Educational Governance*, ed. H. D. Meyer and A. Benavot. Oxford: Symposium.
- Ministry of Education and Training. 2008. *Report on the Situation of Academic Learning Activities of Children*. [In Japanese.] Tokyo: Monbukagakusho Hokokusho.
- Mori, I., and D. Baker. 2010. "The Origin of Universal Shadow Education: What the Supplemental Education Phenomenon Tells Us about the Postmodern Institution of Education." *Asia Pacific Education Review* 11 (1): 36–48.
- Mullis, I. V. S., M. O. Martin, T. A. Smith, R. A. Garden, K. D. Gregory, E. J. Gonzales, S. D. Chrostowski, and K. M. O'Connor. 2003. *TIMSS Assessment Frameworks and Specifications, 2003*. Chestnut Hill: International Study Center, Lynch School of Education, Boston College. [http://timssandpirls.bc.edu/timss2003i/PDF/t03\\_af\\_book.pdf](http://timssandpirls.bc.edu/timss2003i/PDF/t03_af_book.pdf).
- OECD. 2009. *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. Paris: OECD.
- OECD. 2011. *Quality Time for Students, Learning in and out of School*. Paris: OECD.
- OECD. 2012. *Strengthening Integrity and Fighting Corruption in Education: Serbia*. Paris: OECD.
- Oller, A.-C., and D. Glasman. 2013. "Education as a Market in France: Forms and Stakes of Private Tutoring." In *Private Tutoring across the Mediterranean: Power Dynamics and Implications for Learning and Equity*, ed. M. Bray, A. E. Mazawi, and R. G. Sultana. Rotterdam: Sense.
- Park, H. 2013. *Re-evaluating Education in Japan and Korea: Demystifying Stereotypes*. New York: Routledge.
- Paviot, L. 2010. "Paid Tuition: A Potential Challenge to EFA." *IIEP Newsletter* 28 (3): 9.

- Pereyra, M. A., H. G. Kotthoff, and R. Cowen, eds. 2011. *PISA under Examination: Changing Knowledge, Changing Tests, and Changing Schools*. Rotterdam: Sense.
- Robitaille, D. F., and A. E. Beaton. 2002a. "A Look Back at TIMSS: What Have We Learned about International Studies?" In *Secondary Analysis of the TIMSS Data*, ed. D. F. Robitaille and A. E. Beaton. Dordrecht: Kluwer.
- Robitaille, D. F., and A. E. Beaton. 2002b. "TIMSS: A Brief Overview of the Study." In *Secondary Analysis of the TIMSS Data*, ed. D. F. Robitaille and A. E. Beaton. Dordrecht: Kluwer.
- Runte-Geidel, A. 2013. "La incidencia de las clases particulares en España a través de los datos de PISA" [The incidence of private tuition in Spain through the PISA data]. *Revista Española de Educación Comparada* 21:249–82.
- Rutkowski, L., Eugenio Gonzalez, M. Joncas, and M. von Davier. 2010. "International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting." *Educational Researcher* 39 (2): 142–51.
- Silova, I., V. Büdienė, and M. Bray, eds. 2006. *Education in a Hidden Marketplace: Monitoring of Private Tutoring*. New York: Open Society Institute.
- Song, K. O., H. J. Park, and K. A. Sang. 2013. "A Cross-National Analysis of the Student- and School-Level Factors Affecting the Demand for Private Tutoring." *Asia Pacific Education Review* 14 (2): 125–39.
- Southgate, D. E. 2009. "Determinants of Shadow Education: A Cross-National Analysis." PhD diss., Ohio State University.
- Stevenson, D. L., and D. P. Baker. 1992. "Shadow Education and Allocation in Formal Schooling: Transition to University in Japan." *American Journal of Sociology* 97 (6): 1639–57.
- Wolf, R. M. 2002. "Extra-school Instruction in Mathematics and Science." In *Secondary Analysis of the TIMSS Data*, ed. D. F. Robitaille and A. E. Beaton. Dordrecht: Kluwer.
- Zhou, Y., and D. Wang. 2013. "The Family Effect on Extra Lessons in Greater China: A Comparison among Shanghai, Taiwan, Hong Kong, and Macao." Presentation for seminar hosted by the Comparative Education Research Centre, University of Hong Kong, May 22.