PROCEEDINGS 13 1–11
March 2011

The Globalization of Student Assessments and Its Impact on Educational Policy

Takashi HAMANO
Ochanomizu University

Abstract

International student assessments such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) have been widely conducted in recent years, and educational policies tend to be made based on the results of those assessments. However, in order to make policies effective, it is important to garner a deep understanding of international student assessment.

First, it is necessary to know what each assessment measures, and to be aware of any methodological issues therein. For example, often there are problems relating to linguistic and cultural bias, as well as problems vis-à-vis the style of test questions. Sampling and the nature of the population examined are also methodological considerations. Second, in terms of rankings, it is important to pay attention to the number of countries participating, sampling error, and variance. It cannot be categorically stated that "a drop in ranking" in an academic assessment is tantamount to "a drop in academic achievement." A ranking is only ever an expression of a position in relation to other countries, and it changes according to the results of other countries. The more countries that participate, the more difficult it is to achieve a high ranking. In addition, there is some degree of range in both the score and the ranking, due to sampling error. Third, regarding low motivation among Japanese students, it is common in many developed countries for many students to go on to higher education; in addition, there are certain related factors unique to Japan. Fourth, analysis from the viewpoint of gender may be useful for policy-makers and teachers.

International student assessments, although they bear some limitations, can provide useful and important information. It is therefore necessary to provide policy-makers and teachers with feedback that is based on effective research and analysis.

Key words: PISA, TIMSS, OECD, international student assessment, educational policy

1. Introduction

In December 2007, the results of the third PISA[1] (hereinafter "PISA 2006") were released. PISA was conducted for the first time in 2000 (i.e., PISA 2000) and for the second time in 2003 (i.e., PISA 2003), making this the third time that such results had been released. Around 400,000 15-year-olds from 57 countries and regions took part in PISA 2006. The number of countries and regions participating had increased from 32 countries in PISA 2000 to 41 countries/regions in PISA 2003. This increase in participation suggests a growing international interest in PISA.

PISA results bear a large influence on the educational policies of participating countries.[2] For example, in Germany, "PISA shock" inspired various education reforms. In Japan, too, various policies aiming to increase academic achievement have been introduced. For example, when PISA 2003 indicated a drop in student performance, Japan's Ministry of Education, Culture, Sports, Science, and Technology announced the "improved reading abilities program"; furthermore, after the PISA 2006 results were released, a policy to increase teaching time for science/mathematics and the Japanese language was included in the new curriculum guidelines.

These reforms are meaningful in themselves; however, to make them more effective, it is necessary to investigate the various student assessments in greater detail and gain a deeper understanding of them. Although international student assessments such as PISA and TIMSS[3] have limitations, it is possible to gain from them valuable insights that cannot be gleaned from domestic assessments. This paper will focus on international student assessments such as PISA and TIMSS, and discuss their special features; it aims to determine not only in what ways their results should examined, but also how analysis and comparisons should generally be undertaken.

2. Characteristics and methodologies of international student assessments

(1) Special characteristics of international student assessments

Although we usually refer to "international student assessments" as a single, collective item, several different types of assessment are presently being carried out. The main assessments are listed in Table 1. They include International Association for the Evaluation of Educational Achievement (IEA) assessments (TIMSS and PIRLS[4]), as well as PISA types that do not limit the geographic or economic levels of participants; also included are SACMEQ,[5] PASEC,[6] LLECE,[7] and PILL,[8] each of which has limitations *vis-à-vis* the area, language, and the like of participating students. The number of countries participating in assessments with such limitations is relatively low, but these assessments bear the advantage that they can be culturally tailored to the students' knowledge and skills.

Of the assessments listed in Table 1, Japan participates in TIMSS and PISA. Both these assessments include science and mathematics components, but their actual contents differ substantially. PISA tests "how well students can apply their knowledge and skills to real-life problems," while TIMSS tests "how much knowledge and skills the students have attained from the school curriculum."

Table 1: Main international student assessments

Assessment	Years of assessment	Grade or age of students	Content
PISA	2000, 2003, 2006	15 years old	Reading literacy, mathematical literacy, scientific literacy, problem solving (from 2003)
TIMSS	1995, 1999, 2003	4th year and 8th year of schooling (in Japan Grade 4 of Elementary School, and Grade 2 of Junior High School)	Mathematics, science
PIRLS	2001, 2006	4th year of schooling	Reading
SACMEQ	1998, 2000	6th year of schooling	Reading (English), mathematics
PASEC	1995–2001	2nd and 5th year of schooling	Reading (French), mathematics
LLECE	1997, 1999	3rd and 4th year of schooling	Language literacy, mathematics
PILL	1994, 1998–2000	4th and 6th year of schooling	Language literacy, mathematics

Source: compiled from UNESCO (2006).

Thus, PISA measures one's "ability to apply" knowledge and skills, rather than how much of a particular school curriculum one has mastered. This is why the contents covered in the PISA assessment are not called "science" and "mathematics," but "scientific literacy" and "mathematical literacy." PISA is generally known as an "international student assessment," but to be precise, it is really an international assessment of student "literacy." "Literacy" primarily refers to "reading and writing ability," but in recent years, it has also come to mean "basic knowledge and skills in a particular field," as in "computer literacy" or "media literacy." PISA has taken this concept of "literacy" even further by using it to refer not only to "basic knowledge and skills," but also the "ability to apply [basic knowledge and skills]." Recently, even in international organizations like UNESCO, "literacy" does not exclusively refer to the ability to read and write; it has instead become closely associated with "life skills" (UNESCO 2005).

Thus, PISA attempts to measure "application ability," not simply "knowledge and skills." The implication of this is that it is not appropriate to disclaim Japanese educational practices or look for weaknesses in Japanese schooling when the PISA marks or country ranking falls. For example, Japanese students up to the compulsory-schooling stage spend considerable time learning *kanji* (Chinese characters that are also used in Japanese writing) and memorizing English vocabulary, but PISA does not measure these abilities. In this context, to exaggerate the PISA results and consider them a "culmination" of compulsory education is not appropriate.

On the other hand, the TIMSS assessment is for fourth-grade elementary school students and second-grade junior high school students. It tests only mathematics and science subjects, but it also attempts to measure how much school-taught knowledge the student has properly grasped. Thus, PISA and TIMSS test for different abilities, and if the rankings of these two assessments are compared, the results do not always correlate. Countries at the top of the PISA rankings may not necessarily be at the top of the TIMSS rankings. For example, Finland ranked first out of 41 countries in the PISA 2003 scientific literacy section; in the TIMSS 1999 assessment, however, it ranked eighth out of 38 countries. There are many such examples where the rankings of the two assessments do not match, but such discrepancies are largely due to the characteristics of each of the tests. Thus, it is not appropriate to take the results of a particular test and claim that they alone represent the educational level of the country, or that the educational practices of a particular country should be indiscriminately studied or emulated.

(2) "Key competencies" and "zest for living"

PISA is based on the concept of "key competencies," as put forward by the Organisation for Economic Co-operation and Development (OECD). The concept of "key competencies" has much in common with the present aim of Japanese education—namely, to cultivate a "zest for living"—and it is rich with suggestions. However, "literacy" as measured by PISA is one element of "key competencies." The main structural elements of "key competencies" are: (1) the ability to apply socio-cultural and technical tools interactively (i.e., the ability to use language, symbols, and text interactively; the ability to use knowledge and information interactively; and the ability to use technology interactively); (2) the ability to form human relations in diverse social groups (i.e., the ability to create good relationships with others, and the ability to cooperate and the ability to resolve disputes); and (3) the ability to act autonomously (i.e., the ability to be active in the "larger picture," the ability to build and execute a life plan and individual activities, and the ability to express one's own rights, interests, limitations, and needs). However, it is necessary to understand that what PISA measures is just one part of component (1) above. Thus, discourse claiming that school education should set its sights on improving the country's PISA score is misguided: obviously, a test is merely a way of measuring an individual's ability, and it does not measure the solidarity or communality of the citizen.

(3) OECD as an economic organization

One important way of examining international student assessments is through the characteristics of the organization that is leading it. As is well known, the OECD is in charge of PISA and, as its name suggests, it is an international cooperation organization concerned with economic policy. According to Ravi and Lingard (2006), in the early days following the establishment of the OECD, there was a tension between the market liberalism of the United States and the social democracy of Europe. However, in recent times, such ideological debates have diminished, and technical concerns *vis-à-vis* how to most efficiently promote neo-liberalism—with its emphasis on free trade and competition—have predominated. This has strengthened the tendency of the OECD to see education as a means to economic growth. Recently, in this context, dealing with globalization and the intensification of the knowledge economy has become a more predominant focus. There is the common criticism that the OECD, in emphasizing education in relation to the intensification of knowledge economy, is focusing consistently on the economic output of education rather than on its social or cultural aspects (Ravi and Lingard 2006).

In this way, there is a strong tendency for the OECD, in keeping with the character of the organization, to see education in terms of economics. The World Bank has a strong influence over the educational policies of developing countries (Hamano 2005); considering the OECD's view of education as a means of creating human capital and driving economic growth—and, on account of a strong American influence, as a means of advancing neo-liberal policies—the OECD has much in common with the World Bank. However, one difference is that the World Bank has influence over the educational policies of developing countries, on account of its provision of financial loans to those countries; the OECD, On the other hand, does not provide such financing.

In the case of developing countries, it is not just the World Bank that is influential. There is a tension between economic organizations such as the World Bank and several other international organizations: UNICEF, which emphasizes a "human rights-based approach"; the International Institute for Educational Planning (IIEP), which emphasizes capacity development for planning and managing education systems; and nongovernmental organizations (NGOs), many of which emphasize human rights, peace, and social integration. There is a delicate balance to be struck between the World Bank and each of these international organizations. On the whole, however, the organization with the greatest influence on the educational policies of developed countries is the OECD. Most UNESCO activities focus on developing countries, while the influence of OECD on developed countries is strong. It is essential that each developed country critically consider whether it should view education exclusively from an economic standpoint.

(4) Issues with assessment methodologies

The following can be considered methodological problems inherent in international student assessments. First, there is the problem of linguistic and cultural bias. The PISA test is formulated through the following process: a sample list of questions from each participating country is collected and then, after these samples have been duly considered, an international version of the test, in English and French, is set. After a preliminary assessment, the actual assessment test is finalized. As part of this process, issues of cultural bias and the like are carefully debated before the final questions are set. The PISA test questions, including those found in the reading literacy section, are basically constructed in English and French; all other versions are translations. While it is claimed that "cultural bias has been taken into consideration," this is not synonymous with "there is no cultural bias." PISA is meant to assess how well 15-year-olds from each country can apply their knowledge and skills to problems faced in various real-life situations;

however, "real-life situations" differ among countries, and with more and more developing countries taking part, the meaning of "real-life situations" becomes even richer in diversity. Tsuneyoshi (2006) points out the possibility that cultural bias exists, using the mathematical literacy questions as an example; however, such examples can be found with other subjects as well. Finally, Bonnet's (2002) analysis of PISA questions and results show that France scored the highest marks in all questions that France had submitted for the test.

Second, there is the problem of the style of test questions. PISA includes a mixture of answering styles—including free description and discussion, short answers, and multiple choice—but the degree to which students are accustomed to each style differs by country. PIRLS shows clearly the different characteristics of different countries, with distinctions among the countries that achieve high scores on written answers, those with high scores on multiple-choice answers, and so on.

The third problem involves sampling and the size of a country's population. Among the countries participating in PISA, some have large populations and others have small ones. For PISA 2006, for example, the participating country with the largest population of target age group children was Indonesia (4.24 million), and the country with the smallest was Liechtenstein (422). Besides this enormous difference, the sample sizes vary greatly by country. Even if we were to consider only OECD member countries, there is quite a large difference between Mexico (30,971) and Iceland (3,789). In the case of PISA, sampling is carried out amongst 15-year-olds,[9] but in countries other than Japan, this includes students of various school years. Obviously, among students of different school years, the content of what they have studied will also differ. In addition, education systems vary from country to country, and there will be large discrepancies in terms of the timing of the school term and class hours. Of course, these are unavoidable problems, if assessments are limited to those who are 15 years old; conversely, if the limitation is placed on the school year, then the age of the students who participate will vary. In any case, international assessments are likely to feature these kinds of sampling problems. What is important is that the people interpreting the assessment results are aware of such problems.[10]

3. Examining the results of international student assessments

(1) Rankings

The rankings and Japan's position therein are what the Japanese people pay most attention to and what generates large headlines and reports among the Japanese media. Whether it involves PISA or TIMSS, as long as one of the aims of international student assessments is to allow for international comparisons, it is inevitable that interest in Japan will focus on Japan's level of academic performance in comparison to those of other countries, and how this position changes. However, what is important to understand in such cases is the meaning of the rank-order itself, and that one should not place too much emphasis a country's position in this order.

Among the countries that participate in PISA, Japan's rankings are as follows: scientific literacy: 2 (2000), 2 (2003), and 3 (2006); reading literacy: 8 (2000), 12 (2003), and 12 (2006); and mathematical literacy: 1 (2000), 4 (2003), and 6 (2006). Japan's TIMSS rankings are as follows: junior high mathematics: 3 (1995), 5 (1999), and 5 (2003); and junior high science: 3 (1995), 4 (1999), and 6 (2003). Looking at these figures alone, over the last three times that both PISA and TIMSS have been conducted, it seems that Japan's ranking has not gone up even once; if anything, it appears to be exhibiting a downward trend or that nothing is stopping an apparent "downward spiral." Actually, many media reports and editorials took these changes in rank to mean "a drop in academic achievement," which set off alarm bells.

However, several points should be noted. First, one cannot categorically state that "a drop in ranking" means "a drop in academic achievement." The ranking is only ever an expression of a position in relation to other countries, and it changes according to the results of other countries. Just because a ranking has gone down does not directly mean that academic achievement has also dropped. Conversely, if a ranking goes up, this does not mean that ability has correspondingly improved.

Second, a ranking reported by the media is, in most cases, a ranking among "all participating countries." Whether PISA or TIMSS, the number of countries that participate is different each time. The more countries that participate, the more difficult it is to get a high ranking. In the case of PISA especially, the number of countries has almost doubled, from 32 countries in 2003 to 57 countries in 2006; this point must be taken into consideration. If one examines changes in ranking over a certain time period, then looking at changes in ranking amongst *all* OECD member countries will provide a more accurate picture of reality (Table 2).

Third, the media tends to report the rankings of the participating countries by "looking at the average score" and ordering them from the highest score to the lowest. Obviously, rankings require a single index on which the order is to be based; the average score is an effective index, as it is a value that represents the group. However, for both PISA and TIMSS, because they are both sampling assessments, a sampling error cannot be avoided. For this reason, there is some degree of range in both the score and the rank. For example, take the 2006 scientific literacy test: the correct expression for Japan's average score is 531, with a standard error of 3.4; Japan's rank (taking into account the level of statistical significance) amongst the OECD member countries was between 2 and 5 and, amongst all the countries participating in the assessment, between 3 and 9. However, in most cases, this explanation is not provided in media reportage, and the rank alone becomes the only figure quoted. At times, there is the danger that such partial reporting could cause misunderstandings: after all, if the standard error is taken into account, and one looks at the rankings of the OECD member countries only, one cannot say that Japan's ranking has dropped significantly (see Table 2).

It is important also to pay attention to variances other than those in the average score. In the past, regarding Japan's academic performance compared to those of other countries, it has been said that in addition to its average score being in the world's top class, the small variance was also distinctive (Cummings 1980). However, recently, there has been a stronger tendency to argue that our academic performance has become polarized, and that the overall level is not dropping; rather, it has been suggested that the problem is a disparity in abilities. Variance—that is, variation in academic performance, or at least a change in trends—can be seen in international student assessments. Table 2 shows the Japanese and OECD averages, as well as standard deviations. If we look at the standard deviations, one can see that for PISA 2000 in scientific literacy, reading literacy, and mathematical literacy, the Japanese averages were lower than the overall OECD average standard deviation and that variation was comparatively lower. However, since PISA 2003, in scientific literacy and reading literacy, the Japanese averages are higher than the OECD average standard deviation, and for mathematical literacy, Japan's average is about the same as the OECD average.

Table 2: Changes in PISA scores, rankings, and standard deviation

	PISA 2000	PISA 2003	PISA 2006
Number of participating countries	32	41	57
Scientific literacy			
Japan's score	550	548	531
OECD average	500	500	500
Rank within OECD member countries	2	2	3
Range of rank within OECD member countries	1–2	1–3	2–5
Rank within all participating countries	2	2	6
Standard deviation (Japan)	90	100	109
Standard deviation (OECD average)	100	105	95
Reading literacy			
Japan's score	522	498	498
OECD average	500	494	492
Rank within OECD member countries	8	12	12
Range of rank within OECD member countries	3–10	10–18	9–16
Rank within all participating countries	8	14	15
Standard deviation (Japan)	86	106	102
Standard deviation (OECD average)	100	100	99
Mathematical literacy			
Japan's score	557	534	523
OECD average	500	500	498
Rank within OECD member countries	1	4	6
Range of rank within OECD member countries	1–3	2–7	4–9
Rank within all participating countries	1	6	10
Standard deviation (Japan)	87	110	91
Standard deviation (OECD average)	100	110	92

(2) Low motivation and low interest among Japanese students

Regarding international student assessments, one thing that most commentators, including the media, almost always mention is the "low motivation and interest" of Japanese students. Taking PISA 2006 as an example, many editorials and articles raised alarm bells with comments such as "the percentage of students who think that science is useful and are interested in science is lower than in other countries," "the percentage of students who enjoy learning about science is low," and "the percentage of students who think they would like to work in a science-related job is low."

There is the opinion regarding the results of questionnaire surveys that one should be very careful when conducting international comparisons (Murayama 2006): if the basis of comparison is different, then the slant of the answers will also differ. It is therefore essential to be cautious when considering any comparison of countries, but if motivation and interest among Japanese students really are low, then there are possible reasons for this.

First, there is a structural factor: in developed countries, higher education has become more widespread. The issues of motivation and interest become much easier to understand when viewed from the perspective of "developed or developing country." In PISA 2006, countries whose students showed high levels of motivation and interest for science were mainly developing countries in Asia and Latin America. For example, the "index of general interest in science" showed Colombia at number 1, Kyrgyzstan at 2, and Thailand at 3, followed by Tunisia, Mexico, and Jordan—a lineup of developing countries. Japan was number 47 of 57 countries on this index, but South Korea and Finland—the countries with the highest scores for scientific literacy—came in lower than Japan, at numbers 55 and 57, respectively. As a general trend, the countries that ranked highest in terms of this index were almost exclusively developing countries, while the countries that ranked lowest were almost exclusively developed countries. Similarly, the countries that ranked highest in terms of the "index of enjoyment of science" were Tunisia (1), Kyrgyzstan (2), and Colombia (3); developed countries, as a general rule, ranked the lowest.

From this finding, one can say that low levels of interest and enjoyment *vis-à-vis* science is a phenomenon common to developed countries, not just Japan. Why is it that developing countries have such high figures in these indexes? For a start, in developing countries, not all children are able to go on to higher education. For example, in developed countries, almost 100 percent of the children who are at the age set by PISA to be eligible to take part in the assessment are also attending school; however, in many developing countries, this percentage is much lower. Thus, there is the strong possibility that students from developing countries who take the assessment are the "chosen ones," who are already academically orientated and quite a different type of group from the 15-year-olds found in developed countries, almost all of whom are receiving a school education.

The same can be said of TIMSS questionnaire surveys. In TIMSS 2003, the country with the highest percentage of students to answer that they "strongly agree" that "studying science is fun" was Botswana, followed by Egypt, Tunisia, Ghana, and South Africa—all African countries. The percentage of children who attend middle-level schooling in Africa is still low; the overall enrollment rate for middle-level schools in Sub Saharan Africa is 32 percent (2005). In developing countries—especially those with low per-capita incomes—attending school in itself is a "happy" release from child labor, and being able to study at school is considered a great pleasure. From 1995, the author conducted a survey of education in Ghana and was left with the strong impression that students in Ghana's middle schools "wanted to study" and "were so eager to study they didn't want to do anything else." Between students in developed countries, where going to school is taken for granted and compulsory education is free, and those in developing countries, where parents scrape together school fees from their meager incomes, there is a huge difference in what it means to be "going to school and studying".

In PISA 2006, Japanese students were at the very bottom of the "index of instrumental motivation to learn science," and this is often cited as a problem. In other words, in terms of the compiled index—which contained items such as "I study science because I know it is useful for me" and "it is important to study science because it will be useful for future employment"—Japan ranked at the bottom. In this area, too, developing countries were largely found at the top of the index. The background to this is that in developing countries, achieving good grades in math and science means that the student will have a high final score upon completing middle school, which gives one an enormous advantage in securing good employment. Thus, it is very easy for students in developing countries to recognize the connection between studying math and science at school, and future employment. In comparison, it seems to be more difficult for students in developed countries to see a direct connection between studies and future employment, and so many of them do not score highly in terms of the "index of instrumental motivation to learn science."

Japanese students, in particular, are more interested in proceeding to a good high school upon graduation than finding a good job. In fact, in TIMSS 1999, as a reason for studying math and science, Japan's score for "in order to do the job of my choice" was much lower than the international average, but "in order to get into the high school and university of my choice" was above the international average. This finding indicates that Japanese students see the study of science and mathematics as "an instrument" for getting into higher levels of education, rather than as a conduit to future work or employment.

There may also be reasons for Japanese students' low levels of motivation and interest that are inherently Japanese; they involve the ways in which lessons are conducted, the timing of assessments, and the low level of interest regarding science amongst Japanese adults.

Regarding the first reason—the ways in which lessons are conducted—PISA 2006 shows clearly that, in the case of Japan, a low percentage of students think that they are taking classes that emphasize the connection between observation/experimentation and life. In other words, a low percentage of students feel that "the teacher explains how deeply science is connected to life in society," and "students are expected to think about what conclusions can be drawn from experiments." This result reflects a certain indication of the ways in which students are taught in Japanese schools.

Second, it is essential to consider the connection between low motivation and interest, and the timing of the PISA assessment. In Japan, a student's age corresponds with the school year, so students eligible to take the test are all in the first year of senior high school. Japan is the only participating country where this situation exists. In addition, the period in which the PISA tests are conducted is from June to July, when the students are in the first grade of senior high school; the implication is that, until a few months before the PISA test, these students are all struggling with their high school entrance examinations. For this reason, the timing of the PISA test in Japan—namely, immediately following entrance tests—may have some influence on student motivation and interest.

Third, in Japan, it is not just 15-year-olds who lack an interest in science; adults also have the same tendency. An OECD report released in 1998 provides comparisons of the public's interest in science and technology and its level of knowledge, in 14 countries around the world. Of the 14 countries, the Japanese people's interest in science is remarkably low (Shiomi 2000). This data suggest that children and adults alike have a low interest in science, suggesting that the adults' low interest reflects in the country's children.

(3) Effort levels

In connection with the issue of "motivation," one should also consider the issue of how seriously Japanese students treat the PISA. It is only natural for a simple question to arise—namely, "Do students take the test seriously?"—if the test does not directly affect school marks or entrance examinations. In terms of social-survey methodology, whether or not respondents "give their full attention" and "answer seriously" are extremely important issues, and they do not apply to PISA alone. Of course, there are those who think that being able to take seriously something like a test correlates with, or at least is symptomatic of, learning ability; however, if there is a large difference in the level of seriousness exhibited by two groups of students, it is dangerous to interpret score differences as proof of differences in ability.

With PISA, this level of seriousness is expressed as "effort," which is measured thus: students are asked to imagine "a situation where it is crucial that their efforts are recognized and they must try their very hardest." The level of effort in this situation is the benchmark, and it is rated as "10"; that situation is then compared to "how would they rate this test [the PISA assessment]?" that they rate on a 10-point Likert scale. According to this scale, Japan's "effort level" is low; in fact, it is the lowest of 41 countries (OECD 2007). Obviously, having a low effort level does not, in itself, suggest that "Japanese students actually could have achieved better scores." In Japan's case, "even when imagining that it would be included in school marks," the effort level is still close to the worst of all participating countries. It could be, instead, that effort levels among Japanese students are generally low for the tests.[11]

4. Analytical viewpoint: an international comparison of gender differences

As mentioned above, it is extremely difficult to compare, on an international basis, academic achievement levels and student consciousness; such comparisons are rife with methodological difficulties. However, international comparisons of "structure of distinction" in social groups—such as men versus women, and regional and class disparities—are popular research themes. Of these, gender studies are more suited to comparative research than other areas such as class disparities, for two reasons. First, in any country, about the same number of males and females tend to be included in a sample. Second, there is no internationally recognized uniform standard for creating an index of "class"; depending on how this index is configured, the category to which a student belongs will change. "Male" and "female," on the other hand, are straightforward categories.

Table 3 shows PISA (2000, 2006) and TIMSS (1995, 1999, 2003) scores by gender: countries/regions where males scored significantly higher; countries/regions where females scored significantly higher; countries/regions where there was no significant difference; and the results for Japan. If we look first at the PISA scientific literacy scores, we see that most countries exhibited no significant gender-based difference. Wherever there were significant differences, they occurred in equal measure, with the number of countries reporting higher scores for males being matched by countries reporting higher scores for females; however, with PISA 2006, there were more countries where females scored significantly higher. To this point, TIMSS had showed more countries where males achieved higher scores in science ability; however, the PISA results show a tendency whereby females perform better in terms of scientific literacy. In Japan, there was no significant gender difference.

Next, looking at PISA reading-literacy scores, we see that in 2000 and 2006, females achieved significantly higher scores in all countries, without exception. In Japan, too, females scored 31 grade points higher. This difference is lower than the average for OECD member countries, but matches the overall trend in which females score higher. We see the same results in the PIRLS (i.e., a reading-literacy assessment for elementary school fourth-graders). Only two countries reported a lack of significant gender difference in PIRLS, and in 43 countries, females achieved higher scores; in no country did males out-score females. This finding suggests that in almost all countries, there is a gender difference in reading literacy, right from the fourth grade of elementary school.

The trend for mathematical literacy is the opposite of that for reading literacy, with more countries showing males achieving higher scores. For PISA 2006, in only one country did females score higher (i.e., Qatar); the majority of countries showed significant gender differences, with males scoring higher. However, in PISA 2000, the majority of countries showed no significant gender difference, and in PISA 2006, 21 out of 57 countries showed no significant gender difference. Japan showed no significant gender-based difference in PISA 2000 mathematical literacy scores, but males scored significantly higher in PISA 2006. Similarly, TIMSS showed more countries with males scoring higher in mathematics, but when one looks at TIMSS 2003, it is clear that countries reporting higher scores for males were equally matched by countries reporting higher scores for females, and that gender differences in Japan are also diminishing.

TIMSS results for mathematics and science after 1995 show that the gender gap is gradually closing (or, there were equal numbers of countries where males and females scored higher). In elementary school (TIMSS 2003), the number of countries where females scored higher in science and arithmetic was equal to that of countries where males scored higher; in Japan, too, there was no significant gender difference.

Why do these gender differences vary in terms of subjects, years, and countries? Why do females in almost all countries score higher on the PISA reading-literacy component than males? Further research is required, before such questions can be definitively answered. However, variations in gender-difference patterns between countries and years bear important implications; these results suggest that gender differences are not innate, universal, or unavoidable, and that effective educational practices and policies can reduce the gender gap. To date, international student assessments have not been rigorously analyzed from the perspective of gender, but as the aforementioned analysis shows, there is the potential to glean much useful information through this type of gender analysis, which can subsequently be used in reforming both policy and practice.

Table 3: Gender differences in international student assessment scores

		Countries/regions where males scored significantly higher	Countries/regions where females scored significantly higher	No significant gender difference	Japan's scores	
					Males	Females
PISA scientific literacy	2000	3	3	25	547	554
	2006	8	12	37	533	530
PISA reading literacy	2000	0	31	0	507	537
	2006	0	57	0	483	513
PISA mathematical literacy	2000	14	0	17	561	553
	2006	35	1	21	533	513
TIMSS mathematics (junior high)	1995	5	0	17	585	577
	1999	2	0	20	585	577
	2003	9	9	27	571	569
TIMSS science (junior high)	1995	16	0	6	564	544
	1999	13	0	9	556	543
	2003	27	7	11	557	548
TIMSS mathematics (elementary school)	2003	5	4	16	566	563
TIMSS science (elementary school)	2003	5	4	16	545	542
PIRLS (fourth-grade reading)	2006	0	43	2	–	–

Note: Shaded areas indicate statistically significant differences.

5. The globalization of student assessments and future issues

(1) The global spread of national student assessments

Internationally, in recent years—particularly from the end of the 1990s through the 2000s—there has been a growing trend in which each country conducts national student assessments. In the late 1990s, 58 percent of developed countries held at least one national student assessment, and figure this grew to 80 percent in the 2000s. In spite of the fact that the percentage of developing countries was lower than that of developed countries, by the 2000s, the majority of developing countries were conducting assessments. On a regional basis, since 2000, the percentage of countries conducting assessments rose in all regions of the world; this increase was especially prominent in the east Asian, Oceania, and central and eastern European regions. Two-thirds of the western European and North American regions were already conducting national assessments in the 1990s, and this rate is also rising (UNESCO 2007).

Thus, worldwide, national student assessments have proliferated. Among developed countries, factors such as globalization, the intensification of knowledge economy, and the spread of neo-liberalism can be said to explain this proliferation, at least in part. More specifically, there is the understanding that, as globalization and the intensification of knowledge economy each progress, improvements to an individual's intellectual capacity will improve not only his or her own life, but will also prove indispensable to the development of his or her national society. This type of thinking—namely, that individual capacity development is the foundation of national development—has long been seen in developing countries. However, one reason for the recent rise in interest surrounding academic abilities could be the Dakar Framework for Action, which was adopted by the World Education Forum held in Dakar, Senegal, in 2000. This framework sets out certain goals to be met by 2015—goals such as ensuring that all children complete primary education and that preschool education programs are expanded, in order to achieve "education for all" (EFA). One of these goals is "improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy, and essential life skills." Thus, developing countries have become very aware of how achieving "measurable learning outcomes" plays a part in their respective educational policy goals.

(2) Future issues

The use not only of international student assessments but also of national assessments within each country is spreading globally. Japan is no exception: the "national assessment of student learning" has been executed since 2007.

Large budgets are generally made available for student assessments, whether they are international or national in scope; they are conducted on a large scale, and the results are expected to provide a clear basis for policy making. In the United Kingdom, however, it has been claimed that the media and policy-makers "cherry-pick" from assessments only those results that suit their purposes and agendas, and that detailed cause-and-effect studies need yet to be done (Torrance 2006). In Japan, it is necessary to deepen the debate on analytical methods and the application of student assessments, but not in a way that moves directly from the assessment results to policy debates on how to improve student abilities. It is important to carry out detailed analyses in a persistent, step-by-step manner.[12] This paper is a preparatory first step in undertaking this type of assessment analysis.

If one relies solely on the results of the PISA questionnaire, Japanese children apparently have low motivation and a low effort level and do not enjoy learning. If this were true, however, why has Japan been able to achieve high scores, among participating countries, in mathematics and science, and above-average reading literacy scores? There must be a factor or factors other than motivation and interest that push up Japan's scores. In the case of Japan, it is still thought that entrance examinations play an important role in halting a student's falling academic performance level and, as Ichikawa (1990) points out, there is strength in the Japanese education system that remains to this day. Japanese children have a strong tendency to see the progression to higher education as the main reason for studying, and this tendency "fits" with Japanese academic institutions' singular focus on entrance exams. For these reasons, educational reforms must take place in tandem with those regarding entrance examinations.

Media reportage tends to focus on seemingly "negative" aspects—such as Japan's falling rank or Japanese students' low levels of motivation—but it is important to have a sound understanding of ranking. Also, it is important not only to point out "bad points," but also to show a readiness to look for Japan's "good points," "strengths," and "strong aspects." More often than not, these "good points" and "bad points" are two sides of the same proverbial coin. Emphasizing bad points all the time (e.g., "the rank has dropped," "motivation is low," and "no feeling of enjoyment") could discourage students and teachers alike.

Various student assessments results, including those of international assessments, will continue to be released, but it is essential to report feedback to the "front lines" of the education system—namely, the schools—where education actually "happens." Although there are various constraints on international student assessments, they have the potential to provide important suggestions for Japan's education system. Although less than 50 percent of junior high school science and social studies teachers have heard of PISA, many of them believe that "it is necessary to conduct classes that instruct students in the types of abilities that PISA assesses" (Benesse

Educational Research and Development Center 2006). It would seem that interest is definitely *not* low on Japan's education front lines. Future challenges involve the undertaking of high-quality analysis and ensuring that feedback, both positive and negative, is received by the people on the ground.

Notes

1. The Programme for International Student Assessment (PISA) is managed by the OECD. Unless otherwise specified, all PISA data used in this paper are based on the Japanese-version reports published by the National Institute for Educational Policy Research or on the PISA website (http://www.pisa.oecd.org/).
2. For its academic conference in 2005, the Japan Comparative Education Society conducted research on the topic of "demand for international student assessments in each country and the direction of educational reform."
3. This is an assessment that started in 1995, with the aim of continuously assessing international trends in arithmetic/mathematics education and science education. Unless otherwise specified, TIMSS data used in this paper are culled from the TIMSS website (http://timss.bc.edu/).
4. The Progress in International Reading Literacy Study (PIRLS) is a reading literacy assessment aimed at elementary school fourth graders.
5. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) is a student assessment used in English-speaking countries in Sub-Saharan Africa (mainly eastern and southern Africa).
6. The Programme d'Analyse des Systemes Educatifs de la CONFEMEN (PASEC; French African education system analysis program) is a student assessment in Francophone Africa.
7. The Laboratorio Latino-americano de Evaluación de la Calidad de la Educatión (LLECE; Latin American quality of education research) is a student assessment in Latin American countries.
8. The Pacific Islands Literacy Levels (PILL) is an assessment in Oceania countries that tests literacy and numeracy abilities.
9. To be precise, "students who are aged between 15 years, 3 months and 16 years, 2 months at the time of the assessment and who are attending school."
10. A note must be added with regard to population size. Of the OECD member countries, Japan has the fourth-largest population of children who are in the age group eligible to take the assessment (after the United States, Mexico, and Turkey); in terms of the population of children attending school, Japan is third in size (after the United States and Mexico). All countries with higher scores than Japan in the PISA assessment (e.g., Finland, South Korea, Estonia, Canada, etc.) each had a much smaller number of children in school than Japan.
11. The Japanese-version reports cite some problems with the PISA assessment: the test is held in June or July, which is a very humid time of year and the students thus find it hard to concentrate; and while the majority of students try their hardest to answer the questions, an isolated group of students abandoned efforts in the middle of the test and dozed off.
12. Also, if necessary, a supplementary assessment should be conducted—especially a qualitative assessment—even if the sample number is small. For example, the information that can be gained from a questionnaire-type survey on student "motivation" and "interest" is limited, but if combined with a qualitative survey, it may be possible to pinpoint the state of the actual circumstances more clearly.

References

Benesse Educational Research and Development Center (2006) *Research Report on Lower Secondary Education in Japan*, Tokyo: Benesse Educational Research and Development Center. (Originally in Japanese)

Bonnet, G. (2002) Reflection in a Critical Eye: On the Pitfalls of International Assessment, *Assessment in Education* 9(3), 387–400.

Cummings, W. K. (1980) *Education and Equity in Japan*, Princeton: Princeton University Press.

Hamano, T. (2005) World Bank's Policy on Education, *Annual Bulletin of JASEP* No. 12, 83–92. (Originally in Japanese)

Ichikawa, S. (1990) *Theory and Structure of Educational Reform*, Tokyo: Kyoiku Kaihatsu Kenkyu-jo. (Originally in Japanese)

Murayama, W. (2006) Understanding the Result of PISA, *Education and Academic Achievement in Japan*, Tokyo: Akashi Shoten, 70–91.

OECD (2007) *PISA 2006 Science Competencies for Tomorrow's World*, Paris: OECD.

Ravi, F. and Lingard, B. (2006) Globalization and the Changing Nature of the OECD's Educational Work, Lauder, H., Brown, P., Dillabough, J. Halsey, A. H. (eds.) *Education, Globalization and Social Change*, Oxford University Press, 247–260.

Shiomi, T. (2000) Designing Education for the 21st Century, Tokyo: Hitonaru Shobou. (Originally in Japanese)

Torrance, H. (2006) Globalizing Empiricism: What, If Anything, Can Be Learned from International Comparisons of Educational Achievement? Lauder, H., Brown, P., Dillabough, J., Halsey, A. H. (eds.) *Education, Globalization and Social Change*, Oxford University Press, 824–834.

Tsuneyoshi, R. (2006) Academic Achievement in Japan from the Viewpoint of International Comparison, *Education and Academic Achievement in Japan*, Tokyo: Akashi Shoten, 92–106.

UNESCO (2005) *EFA Global Monitoring Report 2006*. Education for All: Literacy for Life. Paris: UNESCO.

UNESCO (2006) *EFA Global Monitoring Report 2007*. Education for All: Strong Foundations-Early Childhood Care and Education. Paris: UNESCO.

UNESCO (2007) *EFA Global Monitoring Report 2008*. Education for All by 2015. Paris: UNESCO.

Acknowledgement

This paper first appeared in the *Annual Bulletin of JASEP* (No. 15, pp. 21–37, June 2008); it had been edited by the Japan Academic Society for Educational Policy (JASEP) and was translated from Japanese with the permission of the editorial board of the JASEP.

Author Note

Takashi HAMANO
Associate Professor
Graduate School of Humanities and Sciences, Ochanomizu University
E-mail: hamano.takashi@ocha.ac.jp