

Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations

Naoki Kuramoto & Rie Koizumi

To cite this article: Naoki Kuramoto & Rie Koizumi (2018) Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations , Assessment in Education: Principles, Policy & Practice, 25:4, 415-433, DOI: [10.1080/0969594X.2016.1225667](https://doi.org/10.1080/0969594X.2016.1225667)

To link to this article: <https://doi.org/10.1080/0969594X.2016.1225667>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 31 Aug 2016.



Submit your article to this journal [↗](#)



Article views: 2557



View Crossmark data [↗](#)



Citing articles: 5 View citing articles [↗](#)

Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations

Naoki Kuramoto^a and Rie Koizumi^b

^aDivision of Research in Higher Education, Section of Admissions, Institute for Excellence in Higher Education, Tohoku University, Miyagi, Japan; ^bSchool of Medicine, Juntendo University, Chiba, Japan

ABSTRACT

Currently, large-scale testing in Japan faces conflicting requirements derived from principles of education on the one hand, and measurement, on the other. Issues of affective ambivalence towards tests (i.e. test aversion and dependence) are also observed. The seemingly conflicting government discussions regarding the national assessment of academic achievement at primary and middle schools, and reforms to university entrance examinations, are discussed here in terms of these issues.

ARTICLE HISTORY

Received 9 October 2015
Accepted 2 August 2016

KEYWORDS

Japanese test culture; the principle of education; the principle of measurement; academic ability assessment; university entrance examinations

Introduction

Educational assessment covers a range of assessment types, from classroom assessment to large-scale testing, for the various purposes of assessing achievement or proficiency, providing diagnoses and conducting placement and selection (Brennan, 2006). This paper describes the use of large-scale educational assessment in Japan, focusing on two national assessments. Both have powerful impacts on both education and test takers, but each has a distinct purpose. First, we consider the role of the national assessment of academic achievement, administered in primary and middle schools, in debates about educational standards. Second, we discuss university entrance examinations and the reform of the university admissions system. We also provide background information on the education system in Japan, and explain social and educational changes and underlying key factors that affect the two assessments (see Ministry of Education, Culture, Sports, Science & Technology [MEXT], 1980; National Institute for Educational Policy Research [NIER], 2015; Organisation for Economic Co-operation & Development [OECD], 2012, for an overview of the Japanese education system).

General education system

The Japanese education system applies the ‘six-three-three-four system’, comprising six years of primary school (for children aged 6–12 years), three years of middle school (i.e. lower

CONTACT Rie Koizumi  rkoizumi@juntendo.ac.jp

Throughout this text, direct quotations originally in Japanese were translated by the authors.

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

secondary or junior high school, for those aged 13–15 years), three years of high school (i.e. upper secondary or senior high school, for those aged 16–18 years) and four years of university. Nine years of primary and lower secondary school are compulsory.¹ According to the Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2013b), a high percentage of junior high school graduates (98.4% in 2013) attend high school or related schools, whereas approximately half (53.2%) advance to education at the tertiary level.

With regard to entrance examinations, public primary and middle schools generally do not use these. Because high school and university are not compulsory, they administer entrance examinations.

Private education plays a substantial role in higher education. As of 2013, only 1.0% of primary schools, 7.3% of middle schools and 26.5% of high schools are private. In contrast, of the 782 universities in Japan, 77.5% are private, with the remainder comprising 11.0% national and 11.5% other public (prefectural or municipal) universities (MEXT, 2013b).

As explained below, this high proportion of private universities affects university entrance examination practices, making the examination system complex and difficult to prepare for.

The Course of Study and changes in educational environment

All Japanese primary and secondary schools are required to follow the Course of Study. The Course of Study sets out the national curriculum standards, which were established ‘with the aim of enabling all students to receive a certain degree of education regardless of the area of Japan in which they receive education’ (MEXT, 2011a). The Course of Study is drafted by selected school teachers and experts in the field of each target subject, and mandated by the Ministry of Education.² Introduced in 1948, the Course of Study has been revised almost every decade since, and is currently in its ninth version.

Reforms in the education system are indicative of various issues faced by Japan over time (see Arai, 2012; Sasaki, 2008). For example, the original 1948 Course of Study for high schools adopted a credit system whereby students selected their subjects, in order to provide education catering to their individual characteristics. The 1956 Course of Study revision increased the number of compulsory subjects to reduce the imbalance of cultural education among students. However, there followed a rapid increase in the proportion of students advancing to high school by the 1960s. In the early 1970s, when this figure had reached approximately 90%, the Ministry of Education urgently needed to handle issues related to widening participation in high school education. Therefore, the 1973 revision of the Course of Study drastically reduced the number of compulsory subjects. This enabled high school students to select from a diverse range of subjects. The Courses of Study of 1982 through 2003 continued this trend. The 1982 catchphrase, ‘fulfilling education free from pressure’, reflected these changes, whereby study time and academic content at school were reduced and students were offered more choices (OECD, 2012). The current Course of Study was implemented in all schools in either 2012 or 2013, and increased the range of curriculum content, after half a century of reductions, due to strong criticism regarding declining academic standards, as described below.

However, changes in educational policy and society are not the only influences on large-scale educational assessment. We would argue that two other factors have also affected such assessment, namely (a) conflicts between principles of education and measurement and (b) affective ambivalence towards tests. We explain these key concepts below, and

suggest that they have affected government debate and reform, as well as public opinion, concerning large-scale assessment in Japan, despite policy-makers not being fully aware of these underlying factors. We then argue that policy-makers ought to recognise the role of these implicit principles and affective factors in discussing policy in a realistic manner. Such an approach may be a key to creating an effective educational measurement system.

Japanese test culture and the principle of education

Arai and Mayekawa (2005) surveyed nine large-scale tests administered in Japan and extracted six common key features that represent Japanese test culture. The first three of these six characteristics are related to the present paper (Arai & Mayekawa, 2005, p. 82):

- (1) Examinations are administered simultaneously, once a year.
- (2) All questions are new each year.
- (3) The questions are subsequently published.
- (4) Content specialists (e.g. university professors) outside the test organisation develop the questions; no psychometricians are involved in test development.
- (5) Scores are reported as raw scores.
- (6) The average time for answering a question ranges from two to four minutes.

Why are tests with all new questions administered simultaneously and later made available to the public? We argue that this is because test items are primarily considered learning materials and indicate the policies of the particular school by which the test is administered. Therefore, tests are mainly judged according to their usefulness for education. This educational usefulness is a central concept in Japanese test culture, and becomes the implicit principle framing discussion of the existing test system and test reform. In this paper, we refer to this as the ‘principle of education’. This principle of education is based on the idea that tests should be useful to education and should have a positive impact on education, improving students’ motivation, achievement and proficiency.

The principle of measurement

In contrast, in the global field of psychometrics and educational measurement, test quality in large-scale testing is usually judged in terms of how well a test functions as a measurement tool. In this paper, we call this the ‘principle of measurement’, and regard it as standing opposed to the above-mentioned principle of education.

With regard to test quality standards, the American Psychological Association (APA) published standards for test quality in 1954 (APA, 1954), and various test standards and guidelines were later published by other organisations. Japan had no quality control standards of this type until 2001, when the Japan Language Testing Association (JLTA) published the Code of Good Testing Practice (JLTA, 2001; Thrasher, 2004). In 2007, the Japan Association for Research on Testing (JART) published test standards, which take practices in Japan into account and suggest appropriate practices from measurement perspectives. The JLTA and JART aim to advance research and improve assessment practices among relevant specialists and practitioners in Japan. The JLTA targets second/foreign language assessment, whereas the JART focuses on assessment in general. Both associations aim to enhance public assessment literacy and to provide useful information on assessment for

decision-making. However, they are not directly involved in discussions on national assessment of academic ability and university entrance examination reforms.

Returning to the principles of education and measurement defined above, the latter implies that a large item bank of thousands of statistically calibrated test items is indispensable. For purposes of calibration, and to exclude faulty items, all items must have been administered to examinees at some point (Ferrer & Grimm, 2012). Although this system has certain advantages, conflicts arise between the principle of education and principle of measurement, mainly with respect to two issues. The first issue concerns the confidentiality of items following test administration (JART, 2007). If items are to be used as learning materials, they must be made available to the public after the test. However, this makes it virtually impossible to develop and maintain a secure, extensive and well-functioning item bank. The second issue concerns desirable item formats (JART, 2007). Test items should ideally provide appropriate measures of complex cognitive abilities (e.g. the ability to express ideas through writing and to think critically), and item formats enabling the assessment of such abilities tend to be robust against cheating or test-taking techniques. However, such items are difficult to administer in large-scale standardised tests because they require long testing times and scoring by human raters.

The test standards of the JART (2007) discuss human scoring extensively and specify the conditions under which test items should be disclosed. Unfortunately, these test standards and the principle of measurement are not well recognised by the Japanese public, as evidenced by their rare mention in public debate. A lack of understanding of testing standards and the principle of measurement leads to less desirable practices in test construction, administration and use. Examples include the vague description of what a test is intended to measure and the difficulty interpreting what test items actually do measure. These shortcomings are reflected in national assessments of academic achievement (Kuramoto, 2011; Okabe, 2011), as mentioned again below.

Affective ambivalence: test aversion and dependence

In addition to the two principles discussed above, affective ambivalence towards tests in Japan has affected public debate. It has long been said that students in Japan cannot help being averse to tests due to ‘degreeocracy’, in a society that heavily emphasises the attainment of an educational degree (Amano, 1983). University entrance is considered one of the strongest determining factors in social stratification, as suggested by Galtung (1971): ‘in a *degreeocracy social birth takes place later than biological birth*’ (italics original, p. 139). In this context, tests are not only understood to be indicators of educational achievement, but also learning materials that force students to study for entrance examinations. This use of tests has led to students’ test aversion, as reported by Yamagishi (1991), who studied tests in contemporary Japanese school education. Yamagishi listed the negative consequences of testing as ‘[students] being labelled, [and testing] promoting a competitive atmosphere in the classroom’ (p. 181). She even argued that ‘belief in those negative consequences of testing is so widespread among the Japanese public and mass media that almost every problem with Japanese youth is attributed to testing’ (p. 181). Similarly, Sasaki (2008) also mentioned students’ ‘extreme anxiety’ and even suicide as a result of highly competitive university entrance examinations (p. 70). Such negative feelings about the use of tests are considered symptoms of test aversion.

With regard to a comparison between Japan and other countries, it is not possible to determine whether Japanese school education is more biased towards testing, or whether Japanese students have greater test aversion, as little cross-cultural comparative research has been conducted. According to Kariya (2002), even in the days when the harsh competition to enter university was labelled 'examination hell', applicants reported getting sufficient sleep and feeling that studying for the examinations was fulfilling. However, policy-makers have tended to believe that test aversion is prevalent among Japanese students, as reflected in the report of the Central Council for Education (1997). Thus, test policy discussions focus on ways to revise testing systems to ameliorate test aversion.

Along with test aversion, we would argue that a complementary type of affective ambivalence can also be observed, namely a dependence on tests. In terms of such dependence, we have observed that Japanese people tend to embrace and commend the quality and impact of tests, particularly those administered by internationally acclaimed organisations. Such people appear to believe in test results with little scrutiny. We propose that people who depend on tests in this way tend to believe that changes in tests and test systems will automatically lead to improvements in education (see MEXT, 2014a). This thought pattern appears to be based on the above-mentioned belief that entrance examinations determine subsequent social status (Galtung, 1971). Test dependence can be observed in official reports and reactions towards international test results.

The two aspects of affective ambivalence discussed above, along with the principle of education and the principle of measurement, can be seen to have affected discussions on large-scale educational assessment in Japan. Against the background of these key concepts, we proceed now to cover two examples of large-scale educational assessment in Japan, each of which has its own focus but both of which have a significant impact on education, namely (a) the national assessment of academic achievement and (b) university entrance examinations.

National assessment of academic achievement and related issues

Issues around declining academic achievement

After the Second World War, the Japanese education sector had two major debates regarding the overall decline in academic achievement, which resulted in the introduction of two national assessments of academic achievement. The first debate was sparked by criticisms on the first Course of Study, which was introduced in 1948 (see above). This Course of Study strongly focused on an education system based on empiricism. It was criticised for causing a decline in academic achievement. This perceived achievement decline led to the introduction of a national assessment of academic achievement. The Ministry of Education conducted an academic survey among samples of students drawn from across the nation in 1956, and a complete survey of all middle school students in 1961.

The above-mentioned post-war national assessment of academic ability continued in use until 1964, when it was discontinued during a period of political conflict. At the time, the Japan Teachers' Union had significant power, and opposed many policies of the Ministry of Education, arguing that the Ministry imposed centralism. The Union was concerned that such national assessment was used by the government to control educational content and force a poor educational curriculum upon students (Adachi, 2002; Schoppa, 1991).

A second debate on the decline in academic achievement began at the end of the twentieth century and continues to this day. This debate began with the publication of a book with the shocking title *University students who cannot perform calculations using fractions* (Okabe, Tose, & Nishimura, 1999). Until then, Japanese people were apparently hesitant to mention individual differences in ability, and issues of academic achievement were not commonly discussed. Kariya (1995) regarded this tendency as an educational mindset considering merit-based views as discriminatory. The above-mentioned book swept such discussion of academic achievement into the public arena.

The Curriculum Council, the advisory body of the Ministry of Education at the time, was tasked with exploring the possibility of a comprehensive survey of academic achievement. The aim was to monitor the achievement of goals and the content of the Course of Study in an objective and ongoing manner. The council proposed a comprehensive nationwide survey of academic achievement among a representative stratified sample of students across school years and subjects for each administration (Curriculum Council, 2000).

Later, the point of contention shifted from a longitudinal decline in academic achievement to a decline in Japan's international rankings. This shift was sparked by the publication of the results of international assessments such as the Programme for International Student Assessment (PISA) and Trends in International Mathematical and Science Study (TIMSS).

Since 2000, the Organisation for Economic Co-operation and Development (OECD) has conducted the PISA every three years, providing international mean rankings in subjects such as mathematics, first language (L1) reading and science (OECD, 2012). Japan's rankings, lower than they had been in the past, led to a sense of crisis regarding academic achievement. The PISA rankings also led to an admiration for the Finnish education system, as Finland garnered a top position until 2009. Many Japanese educational researchers visited Finland to learn their educational philosophy and techniques, publishing books to disseminate knowledge on Finnish methods (e.g. Fukuda, 2006; Shoui & Nakajima, 2005). The Japanese Government appeared to believe that education enabling students to achieve high PISA scores was ideal, without critically examining test content or format (see MEXT, 2014a). Such trends indicate the public's dependence on the results of the 'internationally established tests'.

However, Japanese public opinion appears to apply a double standard. When Shanghai achieved a top position in the 2009 PISA rankings (OECD, 2012), few, if any, Japanese researchers visited the area or acclaimed Shanghai education. One reason may have been that researchers believed this would lead to few insights, as China and Japan share similar testing and learning cultures, with competitive entrance examinations and candidates studying fiercely for success.

In response to the results of internationally administered tests, the Council on Economic and Fiscal Policy (2005) discussed the enhancement of academic achievement through competitive principles. Specifically, the Council proposed the National Assessment of Academic Ability (NAAA) (Central Council for Education, 2005). In addition to the tests of academic achievement that are the focus for this paper, the NAAA also includes questionnaire items on study conditions. The NAAA was introduced in 2007, following a period of approximately 40 years, since 1964, in which no nationwide assessment had taken place.

The National Assessment of Academic Ability

The NAAA is administered to sixth-year primary and third-year middle school students in three subjects: L1 Japanese, mathematics and science.³ The assessment was originally administered to all target students in 2007–2009. In 2010–2012, it was administered only to a sample of students, primarily because reliable data from across Japan had accumulated, which steadily advanced the investigation into and improvement of education (NIER, 2010). However, as of 2013, the assessment has been administered to all students (MEXT, 2014b), mainly because expert panels called for finer-grained investigations that targeted all students (MEXT, 2011b). The NAAA adheres to the Japanese test culture and the principle of education, in that the same items are administered to all students simultaneously and are made available after the test has been administered.

The purposes of the NAAA are threefold: (a) to examine the academic achievement and status of students nationwide and to investigate and improve educational outcomes, (b) to establish a continuous cycle of investigation and improvement in education and (c) to apply the results directly, for example, to enhance teaching and learning on an individual basis (MEXT, 2014b). It should be noted that these purposes do not include accountability (OECD, 2012).

Three main problems are associated with the NAAA, the first relates to the above-mentioned principle of education and principle of measurement, and the latter two relate only to the principle of measurement. First, the NAAA has conflicting purposes. Purpose (a) above, determining nationwide levels of achievement, methodologically contradicts purpose (c), improving teaching for students. If the primary intention were to determine levels of academic achievement, a sampling survey would suffice, with little need to provide feedback to students. In contrast, the improvement of individual teaching and learning assumes that all students are targeted and requires feedback on each student's performance on individual test items. Furthermore, purpose (b), examining longitudinal change, which requires expertise in measurement methodology, also contradicts purpose (c), providing feedback to students and teachers. In short, purposes (a) and (b) are related to the principle of measurement, while purpose (c) is related to the principle of education. Assessment design ought to differ substantially according to which purpose is focused upon (Newton, 2007).

Second, from the measurement perspective, the construction, administration and interpretation of the NAAA differs from that of the typical test framework used in large-scale testing (e.g. Brennan, 2006; Lane, Raymond, & Haladyna, 2016). For example, although a sufficient number of test items should be administered to defend its interpretation as a measure of nationwide achievement (purpose (a)), the NAAA has only a limited number of items. For example, according to MEXT and NIER (2015), the 2015 science test for third-year middle school students (aged 15 years), addressing basic knowledge and skills, and the ability to apply them, comprised 25 items (10 short, constructed-response questions and 15 multiple-choice questions). The results of this test are interpreted not only in terms of the constructs above, but also in terms of science thinking and expression (assessed by 18 items), skills of observation and experimentation (2 items) and knowledge and understanding of natural phenomena (5 items), based on section scores. Furthermore, each item is interpreted as representing one aspect of the target domain (e.g. the ability to understand wind force based on weather codes; the ability to set up an appropriate task based on an issue). The results are then discussed in terms of the achievement of the relevant aspect, and remedial

measures are recommended (NIER, 2011) and implemented in many regions (e.g. Joetsu City, Niigata, 2015). However, the NAAA contains far fewer items than similar large-scale national or international tests. For example, the 2000 PISA science test for 15-year-old students comprised 35 items, the National Assessment of Educational Progress 165 items, and the TIMSS Repeat 144 items (Nohara, 2001). Despite these limitations, the mean NAAA subject scores in each region are announced annually, and municipal boards of education and schools take the results seriously, and discuss how to remedy any issues, even when differences appear to lie within the margin of error (Kuramoto, 2011). Thus, results based on a small number of items are overgeneralised, interpreted beyond their likely scope of applicability, and discussed in relation to educational outcomes.

A related issue is that purpose (b) of the NAAA is to understand longitudinal changes, but this is not reflected in the test framework, as items are published and the same items not used across test occasions (Ferrer & Grimm, 2012). Only optional items currently remain undisclosed (MEXT & NIER, 2013). To achieve purpose (b), items would need to remain confidential, and a pool of items would need to be developed. This strategy is not well understood or applied in Japan.

The third main problem associated with the NAAA is that the validity of the interpretations and uses based on test scores remains inadequately substantiated. The NAAA item format is largely influenced by that of the PISA, with Tests A and B evaluating students' knowledge and ability to use that knowledge, respectively. The ability to use knowledge is defined as the ability to think, judge and express what knowledge and skills are necessary to solve problems. Since Test B aims to assess the ability to use knowledge, empirical evidence is required to demonstrate that it can predict real-life problem-solving ability. Thus, questions related to test validity ought to be raised, for example, in terms of whether paper-and-pencil tests can appropriately assess such an ability to use knowledge (Kuramoto, 2011). Indeed, all interpretations and uses of tests ought to be validated with test contexts in mind (Kane, 2006). However, possibly because the Japanese public generally appears to believe in the seemingly sophisticated international measures, they remain unaware of any criticisms of the NAAA, and of the need to examine its validity. Thus, the test dependence of the general Japanese public leads them to trust test results without questioning test validity in terms of the degree to which tests measure what they are intended to measure.

University entrance examinations

General examinations in the university admissions system

University entrance examinations in Japan are of four main types: general examinations, recommendation-based examinations, Admissions Office (AO) examinations and special selection examinations. Although the Ministry of Education originally instructed that academic tests should not be used for the latter three forms of examination (Kuramoto, 2009), they have increasingly included academic tests in the selection process. Thus, they are considered forms of entrance examination here, as discussed below in the section titled *Selection without academic tests*.

The tests that candidates are required to take in general examinations vary across school type. At national and public universities, applicants take the National Center Test for University Admissions (hereafter the Center Test), followed by a university-developed

test. The Center Test is administered to approximately 550,000 examinees annually (National Center for University Entrance Examinations, 2015). In contrast, private universities use three types of general examination, requiring candidates to take either (a) only a university-developed test, (b) only the Center Test or (c) both the Center Test and a university-developed test. High school students can apply for up to two national universities and as many private universities as they wish. A complex application process is usually required for each university.

A distinct characteristic of Japan's complex university admissions system is that each university develops and administers its own test (Sawa, 2015). Faculty members are charged with selecting applicants, and each university is responsible for its own selection process.

Tests at national universities are based on a unified examination framework adopted by the Japan Association of National Universities (JANU, n.d.), which is in charge of setting guidelines for entrance examinations. In contrast, each public and private university decides on its own examination framework, reflecting its particular management strategies. Thus, the quality of neither public nor private university-developed tests is examined or assured by external associations. The Japanese public generally believes university entrance examinations to be overly complex, with some students describing the experience of taking entrance examinations as 'examination hell or war' (Takeuchi, 1997, p. 193).

General examinations and the principle of education

Both students and teachers often consult past papers for high school or university entrance examinations when they study or teach, respectively, for such examinations. Some of those external to school education believe that entrance examination preparation requires simply cramming the knowledge to be tested, which is useful only for that particular test, and has no substantial educational significance, as mentioned by MEXT (2013c, 2014a). However, educators appear to believe that entrance examinations connote important educational messages from high schools, and that such examinations are beneficial for learning (Takanashi, 2011).

Within the complex university admissions system, in which each university administers its own test, students aiming for admission to prestigious universities and their teachers, may try to identify a university's typical test characteristics from past tests. For example, a mathematics teacher in one of the most competitive academic-track schools in Nagano compared mathematics tests for the University of Tokyo, Kyoto University, Hitotsubashi University and Tohoku University and interpreted their distinctive features. He reported that

Tohoku University presents test items reflecting the importance of 'national foundations' as a backbone. It emphasises the basics and presents items that require candidates to tenaciously solve complicated questions till the end ... I feel building a nation requires such qualities. (Takanashi, 2011, pp. 189–190)

Of course, it may not be possible to attribute such deep meaning to all university entrance tests, but some universities believe that their tests clearly communicate their expectations to prospective students (see Nakaune, 2011). It is for this reason that universities disclose test items after they have been administered, another reason being that the public tend to expect test items to be useful for education.

Although Japanese primary and secondary curricula are regulated by the national Course of Study, administration authorities commonly believe that university entrance examinations heavily influence educational practices (Arai, 2012; MEXT, 2014a). Tests are generally

believed to affect learning and teaching, and this washback effect has been empirically examined in the field of language assessment (e.g. Cheng, Watanabe, & Curtis, 2004). An example of this effect is the introduction of a second language (L2) English listening component to the Center Test (see Watanabe, 2013). According to Uchida and Otsu (2013), the listening component was introduced in 2006 after the technical difficulty of ensuring fairness and administrative stability was overcome. This policy change was intended to lead to positive washback in response to rapid globalisation, improving students' motivation and learning, as well as instructional methods (see also Sasaki, 2008). However, due to the complex interplay among examinations, teachers, students and other factors, the expected positive washback has not been observed (Yanagawa, 2012).

Changes in the system of university entrance examinations

University entrance examinations have been criticised, and the admissions system has been repeatedly reformed (MEXT, 1995; Sasaki, 2008), as detailed below. Despite the common belief that the admissions system uses academic tests primarily to select candidates, and that competition for admission remains hard-fought, the situation has changed somewhat.

Post-war system of university entrance examinations

Post-war university entrance examinations began with the introduction of an academic aptitude test in 1947. This system depended on American intelligence and aptitude tests, and was not well accepted. Such testing was ceased in 1954, after which each university began to administer its own academic admissions test. With this change, entrance examinations became more difficult, with challenging questions, requiring candidates to prepare for the specific test they would take, thereby skewing higher education by primarily teaching to tests. This probably led to the firm image of test aversion for university entrance examinations mentioned above.

A centrally developed test, the Joint First-Stage Achievement Test (JFSAT), was introduced in 1979 to address some of the issues associated with the locally made university tests, but was limited to national and public university applicants. The candidates took the JFSAT first, followed by each university's second-stage exam. The JFSAT covered five subjects within seven courses based on the Course of Study. It was introduced primarily to avoid cramming for specific examinations. However, it garnered criticism for overburdening students, due to its testing of too many subjects and wide overlap with individual university examinations. It was also thought to skew high school education using only a multiple-choice format, which is said to severely restrict candidates' thinking style, and to make university rankings transparent using the same test⁴ (JANU, 1986).

In 1990, the JFSAT was updated to the Center Test, which has been administered with certain modifications ever since (MEXT, 1995; Sasaki, 2008). The Center Test differs from the JFSAT in two main ways. First, it allows candidates to choose what subject test to take, making university rankings less transparent. Second, candidates can report their test results in private university applications, whereby the test attracts a wider range and larger number of candidates. This has resulted in the test having a greater influence on high school education. Interestingly, the Center Test retains the much-criticised multiple-choice format of the JFSAT. In terms of the principle of measurement, the test properties of the Center Test

have deteriorated, as test users can use their results beyond their measurement properties. For example, the scores of different course tests within the same subject can be compared (e.g. biology, physics and chemistry course tests within science) after score adjustment based on means. Even though this is psychometrically indefensible, universities make comparisons between raw scores on tests of different subject areas (e.g. L1 Japanese, L2 English and Math), and use the aggregated results for selection (Kuramoto, 2013; see Coe (2008) and Lamprianou (2009) for methods and discussions of maintaining test comparability).

In the face of falling numbers of 18-year olds caused by declining birth rates, the 1990s saw increased competition among universities to recruit candidates. This led to changes in university entrance examinations, such as the Center Test and university-developed tests. The Center Test began to allow candidates to select from a list of seven subjects, in what was known as the *à la carte* system. In their own tests, many universities reduced the number of subjects assigned. Additionally, the focus of university entrance examinations shifted from strict selection to recruiting candidates without competitive elimination.

Selection without academic tests

In the same period, during the 1990s, universities were strongly encouraged to adopt admission methods that did not use academic tests, under the banner of the diversification of entrance examinations (Amano, 1992). This led to the development of three types of examination. First, instead of general examinations using the Center Test and/or university-developed tests, universities were encouraged to use recommendation-based examinations. Drawing mainly on letters of recommendation from candidates' high school principals and high school transcripts, universities were instructed by the Ministry of Education to limit the use of academic tests and encouraged to conduct interviews in which academic ability was not assessed (Kuramoto, 2009).

Another type of university entrance examination recommended by the Ministry of Education was AO examinations. These apply processes similar to the recommendation-based examinations described above, but differ primarily in terms of who recommends candidates. In the case of AO examinations, candidates recommend themselves. Each university uses its own discretionary power in terms of selection methods and periods. The use of AO examinations has spread rapidly since 2000, when national and public universities began to apply this method (Kuramoto, 2009).

Japanese university entrance examinations generally assume that applicants are in their final year of secondary school in Japan or have recently graduated. However, candidates with special profiles sometimes apply for university admission using the third type of examination, namely special selection examinations. These examinations target diverse types of applicants, including those with outstanding achievement, returnees from overseas, foreign students and those in the working generation. The documents submitted for this examination type are similar to those for AO examinations. The methods used for these three types of examination overlap, and each university defines its examinations independently.

While national and public universities have high proportions of students enrolled through general examinations with a focus on academic tests, the proportions are lower in private universities. As the majority of universities in Japan are private, nearly 50% of students enter universities without taking academic tests, with academic tests currently playing a limited role in university entrance examinations.

Shift in university entrance examination policies

In 2008, in addition to the criticism of general examinations using academic tests, criticism arose against recommendation-based and AO examination processes. Although intended to promote the consideration of qualities that are difficult to evaluate through academic tests, these procedures drew criticism for undermining academic standards. They were viewed as opportunities for universities to recruit more students by offering a less demanding route to academic courses (Kuramoto, 2009). The Central Council for Education (2008), an advisory board for MEXT, made three proposals in this regard. First, they recommended that recommendation-based and AO examinations be overhauled or heavily modified to assess academic ability more substantially. Second, the use of high school transcripts to evaluate academic achievement was encouraged. Third, a new academic ability assessment was planned, tentatively termed the Articulation (or Connection) Test from High School to University (hereafter the Articulation Test). This academic test was intended to serve multiple purposes, among them the objective assessment of students' academic achievement at high school level (with an eye to improving instructional methods) and to the use of results for university entrance examinations, particularly recommendation-based and AO examinations. The plan was to administer the Articulation Test multiple times a year using computer-based testing. However, the proposal came to be regarded as unrealistic, as it did not take the Japanese test culture into account. Specifically, the principles of education and measurement collided, as the principle of education demands that test items be published after test administration for use as learning materials, whereas the principle of measurement presumes that items are not disclosed, in order that a large item pool may be developed for repeated use.

Recent discussions on university entrance examination reforms

The Center Test of January 2012 revealed its limitations in various ways. The test was flexible enough to allow students to select a diverse combination of subject tests, but this flexibility increased administrative complications, and errors involving test instructions led to numerous applicants having to re-take the examinations.

As of April 2013, the situation developed rapidly. The Headquarters for the Revitalization of Education of Japan's ruling Liberal Democratic Party discussed reforms in English language education and proposed the use of external examinations for English university entrance examinations (MEXT, 2013a). This proposal briefly mentioned general university entrance examination reforms in the context of English education reforms, and triggered extensive discussions on university entrance examinations in general. In October 2013, the Council for Revitalization of Education (MEXT, 2013c) stated that 'the selection method for university entrants' should be transformed 'into one that evaluates capability, motivation and aptitude in a multifaceted and comprehensive manner'. The Council's report proposed introducing tests similar to the Articulation Test, proposed earlier but almost forgotten.

The essence of the Council's proposal may be characterised as ahistorical. For instance, the proposal for university entrance examinations to evaluate abilities, motivation and aptitude 'in a multifaceted and comprehensive manner' is reminiscent of a proposal by the Central Council for Education (1971), which led to discussions of the university admissions systems with an emphasis on high school transcripts. However, the focus later shifted from

transcripts to academic tests, and the JFSAT was introduced. In most current entrance examinations, universities recruit students without harsh elimination processes, using a diverse range of evaluations, including interviews and transcripts. The effective use of transcripts has been an issue since the pre-war era, as they have limitations as measurement tools, due to differences in grading styles across schools. Furthermore, the Council's proposal (MEXT, 2013c) encouraged clear connections between high schools and tertiary institutions, but such connections were already made firm by the proposal of the Provisional Council on Education (1987).

Proposals for radical changes in university entrance examinations

The Special Task Force for High School and University Articulation of the Central Council for Education (MEXT, 2014a) presented a radical proposal in December 2014. Specifically, the Center Test is to be abolished by 2020, replaced by the administration of common examinations of an integrated subject-and-course type and a comprehensive type (to elicit students' ability to use knowledge and skills) using computer-based testing multiple times per year. The individual examinations of each university are to be abolished in principle, and students will be selected based on essays written in L1 Japanese, presentations, group discussions, interviews and other forms of evaluation, including common tests. External examinations assessing L2 English speaking, writing, listening and reading abilities will also be used.

A March 2016 report presented a number of concrete procedures for the reform of high school education and university entrance examinations (Yomiuri Shimbun Kyouikubu, 2016). Two new tests will be introduced in either 2019 or 2020, namely (a) a test evaluating scholastic ability for university entrance applicants (hereafter the Entrance Test) and (b) a high school basic scholastic skill test. The latter is intended to assess the basic academic achievement of first- to third-year high school students, in order to enhance learning and teaching. Since it remains to be seen whether this is to be used for entrance examinations, only the Entrance Test will be detailed below.

The Entrance Test will replace the Center Test, being administered among third-year high school students or older candidates once a year as of the 2020 academic year. It will focus on assessing the ability to think, judge and express ideas and consist of a multiple-choice section with single and multiple answers (for all subjects) and a constructed-response section (for L1 Japanese and mathematics). In the L2 English test, the four abilities of speaking, writing, listening and reading will be assessed, with the speaking section using integrated circuit recorders for the voice recording. A section with open-ended responses will, in 2020–2023, elicit relatively short answers (40–80 Japanese characters) in paper-and-pencil format, and will later be administered on a computer and elicit longer written responses (200–300 characters). This section will initially be scored by humans, but later, will be scored with the help of artificial intelligence systems developed in collaboration with private businesses. Universities will receive different result types across sections, namely test scores and responses to each item in the multiple-choice section and band scores in the constructed-response section. Universities will be required to assess a range of aspects of candidates using high school transcripts, as well as essays, presentations and the Entrance Test. Further details are to be provided by early 2017.

Interestingly, policy-makers in test reform discussions appear to pay more attention to the introduction of constructed-response formats (similar to those of the PISA; MEXT, 2014a) in large-scale university entrance examinations than to several other points in the above proposal (e.g. multiple administrations per year). They appear to consider the current Center Test in a negative light, on the basis of examinations' inadequacy in terms of learning materials, and expect the Entrance Test to lead to substantial positive washback effects on Japanese high school education. In this regard, the limitations of large-scale testing appear to be ignored. In practice, it is difficult to design examinations that function as both learning materials and measurement tools. The backdrop to this lack of insight may be the ambivalent feelings of Japanese people based on a long history of test aversion and dependence.

It may be expected that even if conventional paper-and-pencil-based examinations are eliminated, certain preparatory measures or teaching-to-the-test would occur. New test types assessing integrated subjects and courses and others may mainly assess knowledge, not the ability to use it, and would involve technical and other difficulties in the test construction (see Kuramoto, 2000; Kuramoto & Yanai, 2001). Without considering the principle of measurement, it is difficult to ensure the reasonable validity and reliability of tests, and test administration and test taking may become highly labour intensive. Thus, introducing new test types and systems may not be the ideal solution to current issues in large-scale testing in Japan. Rather, it may be necessary first to fully discuss and clarify the strengths, weaknesses and expected consequences of possible test types and administration methods with the two principles of education and measurement in mind. A next step would be to determine and introduce test types and administration methods that are consistent with test purposes and intended constructs, and that are achievable within the bounds of financial and human costs and resources. Results of the old and new test systems should be comprehensively analysed, and possible improvements should be proposed based on empirical analysis. Such systematic measurement procedures would provide a better solution to address current problems.

Conclusion

We have discussed two large-scale educational assessments in Japan and shown how these have been affected by social and educational changes, as well as conflicts between the principle of education and principle of measurement, along with affective ambivalence towards tests. Regarding the NAAA, the test will probably continue to be administered according to the same structure and procedures. The problems of conflicting purposes, undesirable test construction framework, administration, interpretation and lack of sufficient evidence for validity will thereby continue. With regard to university entrance examinations, concrete procedures for the implementation of the reform discussed above have not yet been presented. However, due to the rapid and drastic changes proposed thus far, it may be predicted that Japanese education may face a certain instability in the near future.

A lack of appreciation for the principle of measurement appears to be related to a lack of testing experts in Japan (Kimura, 2010). Test stakeholders tend to have limited assessment literacy and to lack adequate foundations for accepting test standards and discussing issues based on the principle of measurement. Discussion in the absence of rational analysis is likely to be dominated by emotional principles. It may be necessary for policy-makers to recognise the conflict between the principle of education and principle of measurement,

and test ambivalence, and to initiate rational discussions to ensure improvement in Japan's future educational testing context.

Notes

1. There are some schools that do not follow the 'six-three-three-four system', such as unified middle and high schools and medical, pharmaceutical and dental schools. This paper also does not consider education conducted outside of schools (often termed 'shadow education'), such as cram school education (see Lowe, 2015; Takeuchi, 1997).
2. The Ministry of Education was originally officially named the Ministry of Education, Science, Sports and Culture, and was renamed the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in 2001.
3. Tests for L1 Japanese and mathematics are administered annually, whereas a science test is administered every three years.
4. While making university rankings transparent is considered positive in some countries, it was generally considered to be negative in Japan at this time because it leads to discrimination on the grounds of academic results and capability (Kariya, 1995). However, following the heated debate on the decline in academic achievement, as described in *Issues around declining academic achievement*, this view has changed.

Acknowledgement

We express our gratitude to the editor and two anonymous reviewers of this manuscript for their invaluable and constructive comments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Grant-in-Aid for Scientific Research (C) [grant number 26370737].

Notes on contributors

Naoki Kuramoto is an associate professor at the Institute for Excellence in Higher Education, Tohoku University, Japan. He is a member of the Governing Board of the Japan Association for Research on Testing and the chief editor of the *Japanese Journal for Research on Testing*. He has conducted extensive research on undergraduate admission to universities in Japan.

Rie Koizumi is an associate professor of English at Juntendo University, Japan. She is the secretary general of the Japan Language Testing Association. She has extensively taught undergraduate, post-graduate and teacher-training courses on second language testing and assessment.

References

- Adachi, T. (2002). Gakuryoku tesuto hantai tousou [Battles opposing against the National Assessment of Academic Ability]. In T. Abiko, I. Arai, K. Inaga, I. Iguchi, T. Kihara, K. Kojima, & H. Horiguchi (Eds.), *Gendai gakkou kyouiku daijiten* [Encyclopedia of modern school education] (Vol. 2, pp. 378–379). Tokyo: Gyosei.

- Amano, I. (1983). *Shiken no shakaishi* [Sociology of testing]. Tokyo: University of Tokyo Press.
- Amano, I. (1992). Daigaku nyuugakusha senbatsu ron [Opinion on selection of university entrance applicants]. *IDE (Institute for Development of Higher Education). Gendai no koutou kyouiku* [Modern Higher Education], 338, 5–12.
- American Psychological Association (APA). (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, 201–238.
- Arai, K. (2012). Gakushuu shidou youryou vs. daigaku nyuushi: Sono kattou no kiseki to ima [Course of Study vs. University entrance examinations: Trajectory of conflicts and current situations]. In the Center for Advancement of Higher Education, Tohoku University (Ed.), *Koutou gakkou gakushuu shidou youryou vs. daigaku nyuushi* [High school guidelines for the Course of Study vs. University entrance examinations system]. (pp. 7–37). Miyagi: Tohoku University Press.
- Arai, S., & Mayekawa, S. (2005). The characteristics of large-scale examinations administered by public institutions in Japan: From the viewpoint of standardization. *Japanese Journal for Research on Testing*, 1, 81–92.
- Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger.
- Central Council for Education. (1971). *Kongo niokeru gakkou kyouiku no sougouteki na kakujuu seibi no tameno kihonteki shisaku nitsuite: Toushin* [Basic policy on comprehensive expansion and streamlining toward future school education: Report]. Retrieved from http://www.mext.go.jp/b_menu/shingi/old_chukyo/old_chukyo_index/toushin/1309492.htm
- Central Council for Education. (1997). *Chuuou kyouiku shingikai dainiji toushin no gaiyou* [A summary of the second report by the Central Council for Education]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chuuou/toushin/970605.htm
- Central Council for Education. (2005). *Atarashii jidai no gimu kyouiku wo souzou suru: Toushin* [Creating a new era of compulsory education: Report]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/05102601/all.pdf
- Central Council for Education. (2008). *Gakushi katei kyouiku no kouchiku ni mukete: Chuuou kyouiku shingikai toushin no gaiyou* [A summary of a report by the Central Council for Education ‘Toward the construction of undergraduate course education’]. Retrieved from http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/siryu/attach/1247211.htm
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Coe, R. (2008). Comparability of GCSE examinations in different subjects: An application of the Rasch model. *Oxford Review of Education*, 34, 609–636. doi:10.1080/03054980801970312
- Council on Economic and Fiscal Policy. (2005). *Keizai zaisei unei to kouzou kaikaku ni kansuru kihon houshin* [Basic principles regarding economic fiscal management and structural reforms]. Retrieved from <http://www.kantei.go.jp/jp/singi/keizai/kakugi/050621honebuto.pdf>
- Curriculum Council. (2000). *Jidou seito no gakushuu to kyouiku katei no jishu joukyou no hyouka no ariakta nitsuite: Toushin* [Assessment of students’ learning and implementation of educational curriculum: Report]. Retrieved from http://www.mext.go.jp/b_menu/hakusho/nc/t20001204001/t20001204001.html
- Ferrer, E., & Grimm, K. J. (2012). Issues in collecting longitudinal data. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 275–290). Washington, DC: American Psychological Association.
- Fukuda, S. (2006). *Kyousou yametara gakuryoku sekaiichi: Finlando kyouiku no seikou* [No competition will lead Japan to the world No. 1 in terms of academic ability: Success in Finnish education]. Tokyo: Asahi Shimbun.
- Galtung, J. (1971). Social structure, education structure and life long education: The case of Japan. In OECD (Ed.), *Reviews of national policies for education: Japan* (pp. 131–152). Paris: OECD.
- Japan Association for Research on Testing (JART). (2007). *Tesuto sutandaado: Nihon no tesuto no shourai ni mukete* [Test standard: Toward the future prospect of testing in Japan]. Tokyo: Kaneko Shobou.

- Japan Association of National Universities (JANU). (n.d.). *About JANU*. Retrieved from http://www.janu.jp/eng/about_janu/
- Japan Association of National Universities (JANU), Task Force for Improving Entrance Examinations. (1986). *Kyoutsuu ichiji gakuryoku shiken no arikata wo megutte* [Discussing the future of the Joint First-Stage Achievement Test]. Retrieved from www.janu.jp/pdf/kankou/s611106.pdf
- Japan Language Testing Association (JLTA). (2001). *The JLTA code of good testing practice*. Retrieved from https://jlta.ac/?page_id=35
- Joetsu City, Niigata (2015). *Zenkoku gakuryoku gakushuu joukyou chousa no kekka gaiyou nitsuite* [Summary of results of the National Academic Ability and Situation Assessment]. Retrieved from <http://www.city.joetsu.niigata.jp/soshiki/j-gaku/gakushuujoukyoutyousa-kekkgaiyou.html>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). (pp. 17–64). Westport, CT: American Council on Education and Praeger.
- Kariya, T. (1995). *Taishuu kyouiku shakai no yukue-Gakureki shugi to byoudou shinwa no sengokushi* [Future directions of the mass education society: Post-war history of academism and fairness belief]. Tokyo: Chuko-Shinsho.
- Kariya, T. (2002). *Kyouiku kaikaku no gensou* [Mirage for educational reform]. Tokyo: Chikuma Shobo.
- Kimura, T. (2010). Nihon niokeru tesuto no senmonka wo meguru jinzai yousei joukyou no ryouteki haaku [Quantitative analysis about the professional training of testing in Japan]. *Japanese Journal for Research on Testing*, 6, 29–49.
- Kuramoto, N. (2000). Daigaku nyuushi niokeru kyoutsuushiken no houhouren teki kenkyuu: Kyouka kamoku fukugougata sougou shiken no kouzou [A methodological study for joint university entrance examinations: Structure of cross-subject integrated tests]. *The Japanese Journal of Behaviormetrics*, 27, 81–92. Retrieved from <http://ci.nii.ac.jp/naid/110003812597>
- Kuramoto, N. (2009). AO nyuushi no dokoga mondaika [Problems related to AO examinations: Examining the diversification of entrance examinations]. In Bungei Shunjuu (Ed.), *Nippon no ronten 2009* [Points of contention in 2009 in Japan] (pp. 596–599). Tokyo: Bungei Shunjuu.
- Kuramoto, N. (2011). Hakarerumono, hakarenaimono-Hyouka no genkai wo tou [What we can and cannot assess]. In N. Tose & K. Nishimura (Eds.), *Kyouiku niokeru hyouka to moraru* [Assessment and morals in education] (pp. 143–168). Tokyo: Toshindo.
- Kuramoto, N. (2013). Daigaku nyuushi sentaa shiken niokeru taiouduke no hitsuyousei [On the necessity for linking scores of the Center Test]. *Japanese Journal for Research on Testing*, 9, 129–144.
- Kuramoto, N., & Yanai, H. (2001). Kyouka kamoku fukugougata sougou shaken no mondai naiyou bunseki [Content analysis of cross-subject integrated test items]. *Research Bulletin, The National Center for University Entrance Examination*, 30, 83–108.
- Lamprianou, I. (2009). Comparability of examination standards between subjects: An international perspective. *Oxford Review of Education*, 35, 205–226. doi:10.1080/0305498080264936
- Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2016). *Handbook of test development* (2nd ed.). New York, NY: Routledge.
- Lowe, R. J. (2015). Cram schools in Japan: The need for research. *Language Teacher*, 39, 26–31. Retrieved from <http://jalt-publications.org/tlt/articles/4284-cram-schools-japan-need-research>
- MEXT (Ministry of Education, Culture, Sports, Science & Technology). (1980). *Japan's modern education system: A history of the first hundred years*. Retrieved from http://www.mext.go.jp/b_menu/hakusho/html/others/detail/1317220.htm
- MEXT. (1995). *Remaking universities: Continuing reform of higher education*. Retrieved from http://www.mext.go.jp/b_menu/hakusho/html/hpae199501/hpae199501_2_017.html
- MEXT. (2011a). *Gakushuu shidou youryou towa nanika?* [What is the Course of Study?]. Retrieved from http://www.mext.go.jp/a_menu/shotou/new-cs/idea/1304372.htm
- MEXT. (2011b). *Heisei 23 nendo ikouno zenkokutekina gakuryoku chousa no arikata ni kansuru kentou no matome* [Summary of discussions regarding the National Academic Ability Assessment in Heisei 23 academic year and later]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chousa/shotou/074/toushin/1304351.htm
- MEXT. (2013a). *The second basic plan for the promotion of education*. Retrieved from <http://www.mext.go.jp/english/lawandplan/1355330.htm>

- MEXT. (2013b). *Statistics*. Retrieved from <http://www.mext.go.jp/english/statistics/>
- MEXT. (2013c). *White paper on education, culture, sports, science and technology, special feature 2: Accelerating initiatives aimed at education rebuilding*. Retrieved from http://www.mext.go.jp/b_menu/hakusho/html/hpab201301/detail/1360701.htm
- MEXT. (2014a). *Atarashii jidai ni fusawashii koudai setsuzoku no jitsugen ni muketa koutou gakkou kyouiku, daigaku kyouiku, daigaku nyuugakusha senbatsu no ittaiteki kaikaku nitsuite—Subete no wakamono ga yume ya mokuyou wo mebukase mirai ni hana sakaseru tameni: Toushin* [On integrated reforms in high school and university education and university entrance examination aimed at realizing a high school and university articulation system appropriate for a new era—Creating a future for the realization of the dreams and goals of all young people: Report]. Retrieved from http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1354191.htm (for the abbreviated version in English, see <http://www.mext.go.jp/english/topics/1356088.htm>)
- MEXT. (2014b). *Zenkokutekina gakuryoku chousa (Zenkoku gakuryoku gakushuu joukyou chousa tou)* [National academic surveys such as National Academic Ability and Situation Assessment]. Retrieved from http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/zenkoku/1344101.htm
- MEXT & National Institute for Educational Policy Research (NIER). (2013). *Heisei 25 nendo zenkoku gakuryoku gakushuu joukyou chousa (Kimekomakai chousa): Keinen henka bunseki chousa* [Heisei 25 academic year: Summary of detailed, longitudinal analysis of changes from the National Academic Ability and Situation Assessment]. Retrieved from http://www.nier.go.jp/13chousakekkahoukoku/kannren_chousa/keinen_chousa.htm
- MEXT & NIER. (2015). *Heisei 27 nendo zenkoku gakuryoku gakushuu joukyou chousa houkokusho* [2015 Academic year report from the National Academic Ability and Situation Assessment]. Retrieved from <http://www.nier.go.jp/15chousakekkahoukoku/>
- Nakaune, N. (2011). *Nyuushi mondai wo mochiita koudai renkei—Niigata daigaku vaacharu nyuushi taiken* [Cooperation between high school and higher education using entrance examinations: Niigata University's attempts to offer virtual experiences of examinations]. In the Center for Advancement of Higher Education, Tohoku University (Ed.), *Koudai setsuzoku kankei no paradaimu tenkan to saikouchiku* [Ongoing paradigm shift and reconstruction of the articulation between high schools and universities]. (pp. 65–75). Miyagi: Tohoku University Press.
- National Center for University Entrance Examinations. (2015). *Annual report*. Retrieved from <http://www.dnc.ac.jp/>
- National Institute for Educational Policy Research (NIER). (2010). *Heisei 22 nendo zenkoku gakuryoku gakushuu joukyou chousa: Chousa kekka no pointo* [Heisei 22 Academic year National Academic Ability and Situation Assessment: Major points of assessment results]. Retrieved from http://www.nier.go.jp/10chousakekkahoukoku/10_point.pdf
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14, 149–170. doi:10.1080/09695940701478321
- NIER. (2011). *Zenkoku gakuryoku gakushuu joukyou chousa jugyou aidiarei* [Examples of class activities related to the National Academic Ability and Situation Assessment]. Retrieved from <http://www.nier.go.jp/jugyourei/>
- NIER. (2015). *Education in Japan*. Retrieved from <http://www.nier.go.jp/English/educationjapan/index.html>
- Nohara, D. (2001). *A comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)* (Working Paper No. 2001-07). Washington, DC: U.S. Department of Education. Retrieved from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200107>
- Okabe, T. (2011). *Centaa shiken no arubeki sugata* [Desirable situations of the Center Test]. In N. Taniguchi & K. Yamaguchi (Eds.), *Centaa shiken—Sono gakuryoku ni mirai ha aruka* [Center Test: Is there a future for the academic ability assessed by this test?] (pp. 230–244). Tokyo: Gunjyosha.
- Okabe, T., Tose, N., & Nishimura, K. (1999). *Bunsuu ga dekinai daigakusei* [University students who cannot perform calculations using fractions]. Tokyo: Toyo Keizai.
- Organisation for Economic Co-operation and Development (OECD). (2012). *Lessons from PISA for Japan: Strong performers and successful reformers in education. Author*. Retrieved from <http://dx.doi.org/10.1787/9789264118539-en>

- Provisional Council on Education. (1987). *Kyouiku kaikaku ni kansuru daiichiji toushin* [First report regarding educational reforms]. *Monbu Jihou, Showa 62, August Extra Edition*, 50–75.
- Sasaki, M. (2008). The 150-year history of English language assessment in Japanese education. *Language Testing*, 25, 63–83. doi:10.1177/0265532207083745
- Sawa, T. (2015, April 29). Flawed entrance exam reform [Opinion]. *The Japan Times*. Retrieved from <http://www.japantimes.co.jp/opinion/2015/04/29/commentary/japan-commentary/flawed-entrance-exam-reform/#.VpkG9v14a01>
- Schoppa, L. J. (1991). *Education reform in Japan: A case of immobilist politics*. New York, NY: Routledge.
- Shoui, Y., & Nakajima, H. (2005). *Finlando ni manabu kyouiku to gakuryoku* [Learning education and academic ability from Finland]. Tokyo: Akashi Shuppan.
- Takanashi, M. (2011). Messeiji toshitenou daigaku nyuushi mondai [University entrance examinations that convey messages from universities]. In the Center for Advancement of Higher Education, Tohoku University (Ed.), *Koudai setsuzoku kankei no paradaimu tenkan to saikouchiku* [Ongoing paradigm shift and reconstruction of the articulation between high schools and universities] (pp. 183–198). Miyagi: Tohoku University Press.
- Takeuchi, Y. (1997). The self-activating entrance examination system-Its hidden agenda and its correspondence with the Japanese 'salary man'. *Higher Education*, 34, 183–198. doi:10.1023/A:1003001402176
- Thrasher, R. (2004). The role of a language testing code ethics in the establishment of a code of practice. *Language Assessment Quarterly*, 1, 151–160. doi:10.1080/15434303.2004.9671782
- Uchida, T., & Otsu, T. (2013). Daigaku nyuushi sentaa shiken heno eigo risuningu tesuto no dounyuu ni itaru rekishiteki keii to sono hyouka [The historical background of English listening comprehension tests in the Center Test and their evaluation]. *Japanese Journal for Research on Testing*, 9, 78–84.
- Watanabe, Y. (2013). The National Center Test for University Admissions. *Language Testing*, 30, 565–573. doi:10.1177/0265532213483095
- Yamagishi, M. (1991). Testing in Japan. In K. E. Green (Ed.), *Educational testing: Issues and applications* (pp. 169–195). New York, NY: Garland.
- Yanagawa, K. (2012). *A partial validation of the contextual validity of the Centre listening test in Japan* (Unpublished Ph.D. dissertation). University of Bedfordshire. Retrieved from <http://uobrep.openrepository.com/uobrep/bitstream/10547/267493/1/Yanagawa.pdf>
- Yomiuri Shimbun Kyouikubu (2016). *Daigaku nyuushi kaikaku* [University entrance examination reforms: Reports from Japan and abroad]. Tokyo: Chuokoron-Shinsha.