

Insights in Language Testing:
An Interview with Shozo Kuwata -
A Pioneer of Standardized Rank Scoring in Japan

by Noriko Saitoh and Tim Newfields



Few - if any - persons go through the Japanese educational system without frequently encountering a concept known as *hensachi* (偏差値) – a term that could be translated as “standardized rank score”. Moreover, many parents in Japan feel a keen mixture of anxiety and/or pride regarding the *hensachi* rankings of the schools their children attend. Though it might be hard for young folks to imagine, standard rank scores did not become widespread in Japan until the mid-1960’s. Within a short span of time, this concept infiltrated many secondary and tertiary educational settings to become a de facto measure of scholastic attainment and some even maintain, personal worth. In the Japanese context, *hensachi* signifies far more than a statistical formula – it also represents a pervasive social myth that personal ability can be summed up through a single equation which set school admission decisions.

In this interview, we talk with a person responsible for the widespread adoption of standard rank scores in Japan. Born in 1928 in Nagano prefecture, Shozo Kuwata graduated from what’s now known as Shinshu University in 1950. He worked as a school teacher in the Kanto region of Japan from 1950 to 1963. For 17 years after that he worked at a private educational research institute. This interview was conducted on March 29, 2010 near his home in Yokohama. The original Japanese version of this interview is available at <http://jalt.org/test/PDF/Kuwata-j.pdf>.

How did you become interested in educational evaluation?

I should be clear from the onset that I didn’t attend a teacher training college. In fact, no testing and assessment courses were available during my student days. However, in the process of teaching secondary school I became interested in educational assessment. It seems fair to say that the systematic study of educational assessment in Japan developed mainly after World War II. In fact, when the Allied GHQ-CIE (General Headquarters Civil Information and Education Section) examined our educational practices in 1946, they were surprised by the lack of formalized assessment.

In the early 20th century some of Edward Thorndike's works on educational psychology and measurement were available in Japan. For a brief period in the 1930s it appeared as if his ideas might flower on Japanese soil, but it was not until after World War II that the thoughts of this American psychologist took deep root here. Prior to that, the prevailing trend was for each teacher to make unilateral, absolute decisions about student test performance. Hardly any teachers were interested in educational assessment, and the entire process of evaluating student ability was highly subjective.

Since the post-war reforms, educational assessment in Japan has become more grounded on statistical models, and the use of norm-referenced relative ranking has grown. Most schools have adopted a 3-rank grading system that attempts to rate individual performance in comparison with prevailing standards. This system was widely used only for formal student records and report cards prior to the adoption of the *hensachi* system at the secondary school level. In other words, school placement decisions were still being made partly on the basis of the pre-war evaluative practices. In particular, the use of non-norm-referenced, rank-ordering persisted in the placement decision process.

What prompted you to adopt standardized rank scores?

In 1948 – the second year after I started teaching in Tokyo – a conference on high school guidance was held. During that event, a long table of high school placement aspirations and school exam results for many students were displayed. The teachers in charge of graduate placement for each school generally advised students which high schools to apply to based on their school test scores. One student whom I was responsible for was advised not to apply to the high school of his choice because of a 1-point test score difference. I tried to encourage that student by saying, “Don’t worry – you can do it! Apply to the high school you really aspire to enroll in.” How could I tell him that it was futile merely because of a 1-point exam score difference?

I persistently asked the teachers in charge of school placement for a logical explanation of why that student could not apply to the high school he desired. However, they retorted that their decision was in line with accepted precedents and standard procedures. They asked me to give them a logical explanation for objecting to their decision. This experience made me doubt whether placements were being done correctly. It was then that I realized the necessity of coming up with a statistical rationale for making high-stakes test decisions.

Three years after this, I encountered the works of the so-called "father of modern statistics" - the Belgian statistician, astronomer, and sociologist Lambert Adolphe Quet  let (1796 – 1874). His research on standard deviations was invaluable in helping me understand entrance examination score distributions. After carefully analyzing the distributions of many Japanese high school entrance exam scores - which conformed to a Gaussian curve rather well - I became convinced that Quet  let’s statistical concepts could be used to rationalize placement decisions.

At that time, didn’t the concept of standardized rank score already exist?

Yes, though in Japan it was not widely understood. It can be said that Japan has been about 40 years behind the United States in the field of educational measurement. Educational measurement has a relatively short history in this country. In

the early 20th century the psychologist Lewis Terman (1877 - 1956) did advocate the use of a statistical measure similar to the *hensachi* formula I recommend. Moreover, William McCall (1891–1982) offered some useful insights about how to measure scholastic attainment through t-scores. Around 1920 he was conducting research with Edward Thorndike (1874 - 1949) on cognitive assessment. I suppose that could be considered the dawn of educational measurement.

Standardized rank scoring can be applied to any normally distributed data – information distributed along a Gaussian curve. Scholastic attainment can be analyzed on the basis of such curves, and probability theory gives us a good picture of the likelihood of attaining any given score on a test as well as how persons with that given score stand in relation with their peers. Hence, an objective rationale for making pass-fail decisions about student placement can be obtained on the basis of the bell curve distributions.

Incidentally, the same principles Quet let used in calculating the body mass index (BMI) in 1844 can be used to measure academic performance. As you know, he divided the weight of many individuals by the square of their height to come up with an index of “fatness” or “thinness”. Like many anthropometricists of his day, Quet let cherished the belief that the standard distribution curve could account for many features of what was then termed “social physics”. In the process of researching his works and those of other statisticians, I came to realize that the principles behind the Law of Large Numbers could help me interpret test scores with large data sets such as 10,000, 20,000, 100,000. Analyzing Japanese test data in the light of standardized rank scoring became one of my major life challenges.

It seems that many teachers are opposed to the use of standardized rank scores or that they do not understand this concept correctly. What sort of research on standardized rank scoring have you conducted?

Cram schools, the mass media, and teachers could understand why I advocated standardized rank scores. It has been 50 years since I published a paper on standardized rank scoring and about a decade less since I was dismissed from public school for that advocacy. However, standardized rank scoring has flourished to become a virtual index for university entrance examinations. Standardized rank scoring has gained such a strong foothold because it’s convenient and less prone to error than other scoring methods.

I’ve advocated the use of standardized rank scoring in order to elucidate the meaning of a high-stakes 1-point test score difference. I had no intention of arrogantly criticizing the Japanese educational evaluation system then. My goal was simply to demonstrate a hypothesis that large-scale exam score distributions would closely resemble a Gaussian curve. However, in attempting to explain this 1-point difference logically, I had to address the basic issue of whether entrance exams could really measure student proficiency. If so, the question then arose of how accurate the measure was. The scope of my investigation broadened. For example, I wanted to investigate how to create test questions that were accurate and adequate. I also wanted to explore measurement errors and the question of whether it was actually possible to avoid confounding errors. In the process of researching these issues, I discovered that the mere relative rank of students on a test was an unstable measure. This was simply my private research – it was not intended to shape broad academic or cultural policies. As a consequence, the Ministry of Education (which today is known as the MEXT) was highly critical of standardized rank scoring. They alleged that it was a main cause of scholastic cramming and the worship of test score results.

How have your ideas about educational evaluation changed after recommending the adoption of standardized rank scores in Japan?

Well, the prevailing notion among most teachers and parents at the time I advocated standard rank scoring was that the goal of education should be to produce people who can recite textbook information correctly. Even today many Japanese still adhere to the belief that the rote memorization of factual information is the task of education. However, I have come to realize that there are many aspects of education which cannot be measured through standardized rank scoring. For example, communication skills are an essential – though largely untested - feature of the learning process. Moreover, the passion people have for learning is something no standardized rank scores can reveal.

As standardized rank scores make clear, we have to be more careful about assessment. Every single point must be considered to come up with an accurate standardized rank score. Depending on the input, the mathematical operations yield a completely different number each time. For such reasons I have come to believe that educational placement decisions should not rely entirely upon standardized rank scores. Other factors should also be taken into account. In the process of evaluating each student, of course it is necessary to consider their attitude towards their studies in depth – not just their test scores. Their attitude towards learning, which is often revealed in their faces, is also an important barometer to consider. Evaluation is a process of praising strengths – not just merely weaknesses. To a student, it entails considering whether the things the teacher has said have actually been understood and learned. Therefore, when we consider evaluation we must also reflect on fundamental questions such as, "What is education?" When I ruminate on that question over and over, I feel tempted to reply, "education is the process of communicating hopes and aspirations."

In what other countries are standardized rank scores widely used?

To my knowledge, Japan is the only place where standardized rank scores are a pervasive feature. I’m tempted to say that this is due to Japan’s academic meritocracy. However, there are other highly meritocratic societies such as Taiwan, China, and Korea that have not adopted standardized rank scores on a widespread scale. In those places the ranking of educational institutions appears to depend on the cumulative evaluations of multiple stakeholders such as teachers, parents,

and students. I should point out that *hensachi* ratings in Japan are not conducted by the government, but by the major cram schools. It seems that large cram schools have had a significant impact in shaping the educational future of our youth.

What are your thoughts about the university examination system in Japan?

There's something shameful about the process of how entrance exam materials are selected and then acceptance decisions are made on the basis of raw score results alone by teachers with great prestige at institutions of higher learning. However, the university entrance exam system is essentially determined by the Ministry of Education, Culture, Sports, Science and Technology – individual schools have little choice but to follow its directives. For this reason it seems pointless to criticize the exams of specific schools.

Having said that, I do believe that some university admissions office practices have caused the scholastic attainment of high school students in Japan to dwindle. In particular, in my view the AO exams (a free screening entrance procedure that began in 1990 allowing applicants to bypass standard entrance exams) have led to a lowering in university standards. It seems likely that the AO examination system (which requires a flair for self-promotion) is not suitable to the Japanese “national character” or educational environment. AO exams first started out as an attempt to recruit students with diverse abilities that could not be measured via traditional paper-and-pencil exams. This system originated in the United States and Keio University was the first Japanese institution to adopt it.

We might feel inclined to question an examination system which accepts or rejects candidates on the basis of 1-point score differences, but since universities do not know candidates' actual scholastic ability, perhaps the worry about minor score differences is moot. Universities might regard candidates who got one point higher than others on a given entrance exam as “better students”, even if their actual scholastic ability is in fact low. Conversely, they might disqualify candidates with higher scholastic ability who did slightly less well on a given entrance exam. Universities in general feel no responsibility for this discrepancy.

However, even if standardized rank scoring is employed, we should remember that real student ability can not be measured through an entrance exam. A number of minor factors influence test performance, and it is probably not possible to measure true academic ability through a single exam. Exam scores vary according to the difficulty of questions – however, scholastic ability is a relatively robust trait, so a student who has a standardized rank score of 60 should be able to score 10 standardized points higher than an average student generally in theory. In actuality, however, this rarely seems to happen. One of the reasons has to do with exam measurement error, particularly when dealing with indirect measures of ability such as achievement tests. University candidates therefore need to carefully formulate an examination “game plan” that takes measurement errors into account. According to my research on entrance exam errors, there's about a 60% probability that high school entrance exam scores will fluctuate by about ± 3 standardized points. Strictly speaking, the range of fluctuation varies from student to student. It is therefore impossible for candidates to predict their exam results with precision on the day of their exams. I'm tempted to say that only God knows whether any given exam results will accurately reflect a student's ability at a given time – a certain amount of randomness is inevitable. That is why I've come to regard the entrance exam process as a matter of fate rather than a fully predictable process.

What changes would you like to see in Japan's educational system?

The most basic thing we can do is question the practice of relying on cram schools. From an educational perspective, it should be questioned whether or not they actually nurture children. We should also reflect on whether children ought to acquire a basic learning skills foundation prior to entering elementary school. The common adage that children who cannot study are somehow naturally stupid should be questioned. We should also remember that education involves social networking skills and learning is not merely a matter of putting facts in the head. Perhaps it is important to polish our own innate learning skills. Babies from age 1 to 3 have an inherent ability to process and incorporate new information. Children need to develop fundamental discrimination skills. Parents have the ability to extend the natural endowments of children, and should make it a top priority to cultivate their natural gifts. I like to think of each person as a “mono-culture” that is unique in some ways. Parents can have a significant impact in fostering or hindering their children's development. For this reason I think the proverb, “Children grow up seeing their parents' backs” is apt. Before attempting to reform the Japanese educational system, perhaps we should focus more on parent-child roles. I do not think education in Japan will improve until children are treated differently – not as government employees or servants. Improving Japan's educational system needs to start at the preschool level.

NOTE: When measuring the characteristics of things such as academic ability or height among a group, standardized scores can be obtained by calculating the relative distributions of individual datum away from the mean. Standardized scores are measured in standard deviation units and known as Z-scores. Algebraically, standardized scores are calculated by this formula: $Z = (X - \mu) / SD$ in which X represents the raw scores, μ represents the mean, and SD signifies the standard deviation. Standardized rank scores (*hensachi*) are calculated by a slightly different formula: $10(X - \mu) / SD + 50$. Hence an item with a standardized score of +1.0 (ranked in the upper 15.86% percentile on a normal distribution) would have a standardized rank score of 60. Conversely, a person whose standardized score is -1.0 (ranked in the bottom 15.86% percentile on a normal distribution) would have a standardized rank score of 40. A person scoring right in the middle of a bell curve would have a standardized score of 0 and standardized rank score of 50.

Acknowledgement

Many thanks to Melissa Tsuchiya and Michihiro Hirai for their feedback on this article and on this translation.

Works Cited

- Iwahara, S. (1957). *Kyouiku to Shinri no tame no suikegaku* [Statistics for education and psychology]. Tokyo: Nihon Bunka Kagakusha.
- McCall, W. A. (1922). *How to measure in education*. New York: Macmillan.
- Quetelet, A. (1835, 1842). *Sur l'homme et le developpement de ses facultés, essai d'une physique sociale* [On the development of man and his faculties: Essays on social physics]. Paris: Bachelier, Imprimeur-Libraire.
- Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*. Boston, NY, Chicago, SF: Houghton Mifflin. Available at <http://www.gutenberg.org/etext/20662>
- Thorndike, E. L. (1903). *Educational psychology*. New York: Lemcke & Buechner.
- Thorndike, E. L. (1904). *Introduction to the theory of mental and social measurements*. New York: Science Press.
- Thorndike, E. L. (1927). *The measurement of intelligence*. New York: Teachers College.



Shiken: JALT Testing & Evaluation SIG Newsletter 2010 Editorial Board

Editors and reviewers are approved annually at the SIG Annual General Meeting held during the JALT Annual Conference & Materials Exhibition

Contracted Chief Editor

Tim Newfields (Toyo University Faculty of Economics)
timothy at toyo dot jp

Statistics Column Editor

J.D. Brown (University of Hawaii at Manoa)

Rasch Column Editor

James Sick (International Christian University, Tokyo)

Reviewers

Mark Chapman (University of Michigan, English Language Institute)
Jeff Hubbell (Hosei University)

H.P.L. Molloy (Toyo University Faculty of Business Administration)

Ian Mumby (Hokkai Gakuen University)

Joesph Ring (Toyo University Faculty of Economics)

Ed Schaeffer (Ochanomizu University)

Chris Weaver (Toyo University Faculty of Business Administration)

***NOTE:** If you're interested in serving as a reviewer,
please contact any of the persons listed above*