

A Comparative Analysis of the Japanese University Entrance *Senta Shiken* Based on a 25-Year Gap

Michael (Mike) Guest
University of Miyazaki

This study describes changes that have been made in Japan's National Center Examination for University Admissions (hereinafter *Senta Shiken*) by comparing the 1981 and 2006 versions of the test. An analytical outline of both tests was performed primarily with a top-down focus upon the categories of text type, topic and genre, task type, and skills required. Consideration was also given to the weighting of the various sections and tasks. The purpose of the study was not only to note new developments made in line with recent Japanese Ministry of Education, Culture, and Technology (MEXT) measures regarding university entrance exams and English education policy in general, but also to update the seminal work of Brown and Yamashita's (1995) analysis of Japanese university entrance exams and recommendations later made by Brown (2000). This study also briefly comments on positive washback that the *Senta Shiken* could have on high school English pedagogy in Japan.

本論文は1981年と2006年の大学入試センター試験を比較検討したものである。分析結果を基に、テキストの型、トピックとジャンル、タスクの型、測定対象となっている技能について変化しているのか否かを検証した。検証の主な目的は最近の大学入試の傾向を、英語教育全般に関する文科省の政策に照らして検討するものであった。さらにBrown & Yamashita (1995) による日本の大学入学試験に関するデータを再度検証した。

あわせて、Brown (2000)などで行われている提言の妥当性を再検討することも目的とした。さらにセンター試験が日本の高校英語教育に及ぼす好ましい影響についても簡単に触れた。

Background

Most readers are likely to be familiar with the centrally administered National Center Examination for University Admissions (hereinafter *Senta Shiken*). In the eyes of many this examination stands as the pinnacle of Japan's standardized education system. Each year over 500,000 examinees nationwide sit for this test, one which will have a great impact on determining which university exams, taken some weeks later at a second stage (*Niji*), they will have the best chance of succeeding in. Numbers will vary from individual university to university but many require that candidates obtain a certain score on the *Senta Shiken* before being allowed to sit for the *Niji* exam. Also, the *Senta Shiken* score is factored on a percentage basis (again varying by institution) into the total entry score set by any given university. However, recent changes have affected both the force and function of this exam. As both Mulvey (2001) and Mori (2002) note, Japan's low birthrate has now created a demographic in which there is nearly one seat for every university applicant, so competition is not as fierce nor as all-determining as it once was. Moreover, as mentioned by both Mori (2002) and Arai (1999), this has spawned alternative means of entering universities, such as recommendation systems, local quotas, and so forth. Trelfa (1998) adds that while there is educational standardization at the high school level, greater diversification and power given to local boards of education has weakened the uniformity of this standardization, such that the Japanese Ministry of Education, Culture, and Technology (MEXT) does not closely monitor the national guidelines at the local level, where boards of education have greater authority than ever to determine curricula.

Over the years, MEXT has implemented many changes in the *Senta Shiken* as well as recommendations made for individual university second-stage (*Niji*) exams—the exams which finally determine actual university entry (Monbukagakusho, 2000, 2003). A preliminary common entrance test (widely-known as the *Kyoutsuu Ichiji*) was originally established in 1979 under the leadership of the then Ministry of Education in order to create a unified, national standard that universities could use as a reference point. This was a result of a series of recommendations first made in 1971 by the Board of University Entrance Examination Improvement to alleviate the competition and stress that had hitherto surrounded the individual university entrance exams. In 1977, the National Center for University Entrance Examinations (*Daigaku Nyuushi Center*) was established to design and manage the *Kyoutsuu Ichiji*. One of the main criticisms of *Kyoutsuu Ichiji* was its inability to measure examinee abilities in analysis, creativity, and critical thinking (Kuroha, 1992). Hirezaki (1991)

added that the questions were too difficult, twisted, or obtuse for high school seniors.

Therefore, one of the main goals of the Interim Education Council (*Rinji Kyouiku Shingikai*) in 1984 was the improvement of university entrance examinations. This council was formed separately from the Ministry of Education as a quasi-private consulting body. In the Council's first report, in June, 1985, the establishment of another common test was recommended to replace the *Kyoutsuu Ichiji*. The new common test was proposed with characteristics different from the *Kyoutsuu Ichiji*, such as the improvement of multiple-choice task construction and alternative ways of scoring the exam. The new system began in 1991 under its current name *Daigaku Nyuushi Senta Shiken*.

Several revisions and guideline developments have taken place since. In both 1998 and 2000, MEXT set about reforming the content of the *Senta Shiken*—a process previously undertaken every ten years—with the goal of fostering general comprehension and analytical skills over and above those of memory or recognition. In 2003 the ministry announced its further intention to promote the development of “Japanese with communicative abilities” and recommended that individual university entrance exams try to reflect such a focus in terms of exam form and content. A listening component was established in 2005.

Research on the examination itself is surprisingly scant. References inevitably start with Brown and Yamashita's comprehensive (1995) review and critique of Japanese university entrance English exams, which has remained the seminal English work on the subject, even though the samples on that survey are now over a decade old. Since new approaches and directions have been implemented, one may ask whether these changes and implementations have moved the test away from its critical description, that of a poorly designed, discrete-item-based measure of grammatical minutiae and “testwiseness” made by “amateurs” (see Brown, 2002; Brown & Yamashita, 1995; and McVeigh, 2001), towards a test which comes closer to reflecting (or fostering) healthy pedagogical and educational practices.

Kikuchi (2006) updated Brown and Yamashita's research but focused only upon second-stage *Niji* examinations for prospective English majors at prestigious universities. Using the same categories of analysis as Brown and Yamashita, Kikuchi concluded that little has changed on these selected *Niji* exams, save for the emergence of, “a few new item types, such as summarizing reading passages or listening passages” (p. 90).

However, this study focuses solely on the *Senta Shiken*, since it is the only standardized nationwide English entrance exam and the most heavily weighted. It is developed through government agencies and therefore stands as a bellwether of national policy regarding English pedagogical content (more so than *Niji* exams which may reflect less uniform, more localized concerns and practices).

Ichige (2006) analyzed the 2005 *Senta Shiken*, concluding that “it can hardly be said that the current Center examination measures communicative ability appropriately” (p. 21). The validity of Ichige’s basic research rationale has been questioned by Guest (2007), who argued that the *Senta Shiken* was never meant to be a measure of communicative skills in the first place. Still, Ichige’s study is one of the very few attempts to measure how the *Senta Shiken* has changed in order to reflect new pedagogical trends or emphasis using an item-by-item analysis, a procedure also undertaken in the present study.

Brown (2000) followed up the 1995 research with a series of recommendations for improving the examination system. Among these were comments regarding economic and sociopolitical polity, which falls outside the concerns of this study, but these also included suggestions regarding test construct validity (several solicited from Hughes (1989)) such as:

- using a wide variety of samples,
- testing those abilities that one wants to develop and encourage,
- increasing the variety of examination formats,
- assessing higher order cognitive skills, and
- not limiting texts or tasks to academic fields.

While Guest (2006, 2007) has outlined many of the situational, environmental, and logistic factors limiting the scope of the *Senta Shiken* which render some of Brown’s original recommendations as impractical or implausible, there may well have been positive developments on the exam in the past several years that both reflect and foster healthy pedagogical approaches in the high school system based on such recommendations.

Thus, this research paper’s purpose is twofold. First, it seeks to analyze and compare a recent (2006) edition of the *Senta Shiken* with the 1981 version in order to highlight the developments that have taken place in the intervening years, especially in light of MEXT’s new focus. Secondly,

this paper seeks to confirm the realization of the above recommendations made by Brown (2000) and the subsequent possibility of positive washback onto high school English pedagogy created by changes in the exams instead of negative washback. Negative washback, in terms of fostering a grammar-translation methodology at the high school level, has been noted in past research (Bailey, 1999; Gorsuch, 1998). Mori (2002) further argues that a disjunction between high school and university entrance exam content “has essentially created the area of remedial education” (p. 42).

On the other hand, although many have questioned both the quality and quantity of washback (Guest 2000; Stout, 2003; Mulvey, 2001; Watanabe, 1996), the fact remains that perceptions of both the form and content of the *Senta Shiken*, whether accurate or not, still inform high school pedagogy and policy to a considerable degree.

Research Method and Design

For this study, the analytical template used in Brown and Yamashita’s (1995) study has been avoided. This was a conscious decision made for several reasons. A judgmental rather than an empirical approach has been adopted because it is believed that an empirical approach does not do justice to analyzing or measuring certain important aspects of the *Senta Shiken*, nor is a purely empirical approach entirely sensitive to the environmental constraints that surround the test.

Here I must address the key points of test utility and construct validity. A test is considered valid only when the construct matches the stated purpose of the exam. But what is the purpose of the *Senta Shiken*? None has ever been publicly proposed but there is no doubting that the primary function of the English portion of the *Senta Shiken* is as a type of placement test, to a) determine student aptitude for academic study of English at the tertiary level in Japan, and b) to stratify examinees so that they make suitable choices in terms of deciding which second-stage university examinations to sit for. It is *not* an achievement test or a preparation test for utilizing English skills in real-world communicative settings.

Noting this about test utility and construct brings into question some of Brown and Yamashita’s (1995) foci as well as those of Kikuchi (2006). A primary example can be found in their common concern with test difficulty and the resultant focus on utilizing readability scales. First, test difficulty as a whole can often be largely a matter of task difficulty and the type and number of skills demanded—areas that readability scales are

not able to measure. As a result, this analysis focuses more upon task type and skills required as categories.

Also, the designers of the *Senta Shiken* do not use any vocabulary or other content that is not mandated in the standardized high school curricula (Trelfa, 1998). Trelfa further explains that “scores...cluster nearer and nearer the top with each passing year and it becomes more difficult to discriminate among applicants. In order to distinguish among applicants, the Center Examination has had to create increasingly difficult questions. These questions are still based on the Monbusho curriculum, but involve, in the words of a test preparation manual, “the synthesis of several topics.” In other words, because the *Senta Shiken* is norm-referenced, as long as examinees’ scores can eventually allow them to be adequately stratified for placement in appropriate universities, “difficulty” might be said to be less of a factor in measuring construct validity.

This judgmental approach allows this study to question the categorization of certain discrete-point vs. integrative items. This is because although a question may be posed in multiple-choice format and thereby appear to be of the discrete-point variety, the skill required to complete the task correctly might well demand an integrative approach. Furthermore, there exists a large grey area as to what constitutes discrete-point vs. integration. Knowledge of an individual lexical item may be classifiable as “discrete-point” but knowledge of the function of said item within certain contexts would be “integrative.” It seems that by taking a bottom-up approach in calculating such items both Brown and Yamashita’s and Kikuchi’s studies were thereby predisposed to categorize some integrative items as discrete-point.

This study also takes a *positive* view of utilizing a great variety of text-types, topics and genres, and tasks, a quality criticized by Brown and Yamashita. A wide variety of item types can measure a wider variety of skills, despite Brown and Yamashita’s claims that such a test construct demands greater “testwiseness” on the part of examinees. And, in order to enhance reliability, texts, genres, and topics addressed in a well-rounded English test should not be reduced to single types or patterns which would favor a limited set of skills or particular types of learners.

In short, employing a top-down analysis, this judgmental study aims to focus upon the bigger picture. There is a greater concern for task type and genre, both of which influence the validity of an exam and yet are not widely treated in the background literature. This study also considers the skills demanded of the examinee, as opposed to the narrower concept of

item types (utilized by Brown & Yamashita, 1995), which tends to focus upon the layout of the question, rather than the actual task. The notion of “skills” refers to a variety of cognitive abilities which I believe are more accurate indicators of exactly what the tests are trying to measure. So, noting whether tasks ask examinees to predict, summarize, expound, extrapolate, interpret, infer, sequence, reconstruct, paraphrase, read for specific information, or read between the lines, none of which seem to have been specifically addressed in the previous analytical literature, is of greater interest.

By utilizing this different analytical design, we can gain a very different view of the current *Senta Shiken* in Japan, a perspective that is more sympathetic to the overall test purposes and functions—measuring Japanese students’ aptitudes for academic English in order to stratify candidates into appropriate universities. In taking this approach we do not intend to overturn the research of previous researchers but merely shed a different light on the exams, one which may provide us with a broader view to understand how positive changes have indeed appeared over the years.

The primary analysis of both exams will therefore be based upon four categories: a) text type, b) topic/genre, c) task type, and d) skills required. The scoring weight given to different items or sections, a crucial measure of test priorities, is also always noted.

Analysis of the 2006 *Senta Shiken* English Exam

This analysis of the 2006 exam is based on 2007 Kako Mondaisaku Tanki Kansei Ban available from Kyougakusha, Tokyo. (Note: The weight accorded to each section is represented as “pts.”)

Part 1 Reading: Total 200 pts. (Test total = 250 pts.)

Section 1: 16 pts.

- *Text:* Section a) individual sentences, b) casual conversation of nine turns
- *Topic/Genre:* a) none, b) casual conversation between friends regarding weather and umbrellas
- *Task type:* a) word accent (two questions), b) sentence stress (four questions).

- *Skills required:* a) knowledge of accent patterns, b) understanding of stress in verbal, interactive contexts

Section 2: 38 pts.

- *Text type:* a) various (sentence level), b) short four-turn conversations on various topics, c) various (sentence level)
- *Topic/Genre:* a-c) none consistent or specified
- *Task type:* a) discrete item slot and filler (10 questions), b) conversation slot and filler (three questions), c) sequencing/ordering at the sentence level (three questions double weighted)
- *Skills required:* a) discrete grammatical knowledge, b) norms of social interaction, c) syntax: sentence-level word order, d) specific lexical knowledge

Section 3: 34 pts.

- *Text type:* four paragraphs and one short essay of 150 words (seven questions total)
- *Topic/Genre:* expository magazine article type: image of elephants, throat microphones, gorilla behavior, the cause of allergies
- *Task type:* logical connector (discourse markers) slot and filler, reordering/sequencing
- *Skills required:* understanding rhetorical flow, understanding the role of logical connectors and transition phrases, coherently sequencing information

Section 4: 35 pts.

- *Text type:* 600-word essay plus graph (visual prompt)
- *Topics/Genre:* Expository magazine or journal essay: Attitudes toward volunteer work among different Japanese prefectures
- *Task type:* a) one heavily weighted question asking examinees to interpret an essay via a graph, b) general comprehension, rephrasing (four questions)
- *Skills required:* Applying information to visual prompt, interpreting data, paraphrasing

Section 5: 32 pts

- *Text type:* four-person extended conversation of 24 turns, with two large visual prompts
- *Topics/Genre:* casual conversation (contextualized) regarding description of a car, description of a task, and details on finding the car.
- *Task type:* a) slotting in common interactive phrases, b and c) applying information to two visual prompts, d) true / false
- *Skills required:* applying information to visual prompts, interpreting data, paraphrasing, understanding of conversational set phrases in context, understanding pragmatic force and uptake, displaying general comprehension

Section 6: 45 pts.

- *Text type:* one essay of about 600 words
- *Topics/Genre:* memoir, narrative on the author's change in perception of a neighbor
- *Task type:* a) five general comprehension questions, b) one three-part true / false question
- *Skills required:* holistic reading (general comprehension), summarizing, making inferences.

Part 2 Listening: Total 50 pts. (Test total = 250 pts.)**Section 1: 12 pts.**

- *Text type:* exchanges of two to four turns
- *Topics/Genre:* information on flight times, weather, planning a trip, and service encounters
- *Task type:* listening for specific information (three questions), listening to match visual prompts or illustrations (three questions)
- *Skills:* listening for specific information, making inferences from data, applying information to visual prompts

Section 2: 14 pts.

- *Text type*: seven short exchanges of two or three turns
- *Topics/Genres*: various
- *Task type*: listening in order to predict extended conversational responses
- *Skills required*: understanding of discourse flow, predicting

Section 3: 12 pts.

- *Text type*: a) three four-turn exchanges, b) one 10-turn extended conversation
- *Topics/Genres*: a) instruction, opinion, suggestion b) planning a ceremony
- *Task type*: a) three general comprehension questions, b) applying information to a chart
- *Skills required*: general comprehension, making inferences (pragmatic force and uptake), sequencing data to fit into visual prompt

Section 4: 12 pts.

- *Text type*: a) three short monologues: b) one extended monologue
- *Topics/Genres*: a) short formal speech, answering machine message, school announcement, b) narration on extreme weather experience
- *Task type*: general comprehension questions: a) one question for each of the three short monologues, b) three questions for the extended monologue.
- *Skills required*: a) listening for specific information, making deductions from data, b) holistic listening, paraphrasing

Additional Comments Regarding the 2006 Senta Shiken

Brown and Yamashita (1995) critically observed that *Senta Shiken* questions were all in multiple-choice format. However, the tasks demanded in these multiple-choice questions vary considerably and do not necessarily entail a discrete-point focus. Within a multiple-choice format, there is a

great difference between being asked (a) to correctly identify a discrete point from a text by “recognizing” it among the answers, and (b) choose the correct order of three sentences excerpted from that text as in the following:

#3B 1:

- a. While some people worry that a system like this will be able to read our minds, in its current stage it can understand only a few simple words.
- b. The system is sensitive to pick up the “inner speech” we use when we are silently reading or thinking.
- c. Now, NASA scientists have developed a more advanced system than throat microphones.
Answer (choose the correct order) 1. A-B-C, 2. A-C-B, 3. B-A-C, 4. B-C-A, 5. C-A-B, 6. C-B-A

Likewise, the true / false questions are not simple binary questions that allow readers to skip reading the text and merely scan it for key word(s) in order to choose a T or F answer, but rather involve choosing one to three true or correct inferences, summarizations, or paraphrases from a list of six as in the following example:

#5d. Choose the two of the following that are true based on the text:

1. Owen’s sister has no children, but she keeps a teddy bear in her car.
2. Owen locked his sister’s car because it is not safe to park on the red level.
3. Since Owen’s car needs to be repaired, he borrowed his sister’s car that day.
4. The store where Owen works has plenty of space for employees’ things.
5. Jay, Yuki, and Ella will take Owen’s things to his sister’s car.
6. Owen did not buy very many things while he was shopping with his friends.

The current format for the reading section has been largely in place since 1998, as samples of older exams (*kakomon*) taken from the same source text reveal. *Kakomon* that are representative of the current test

format are usually provided in practice booklets. Very few questions or tasks from pre-1998 are listed in the *kakomon* sections. The various mock exams (*mogi shiken*) provided in practice booklets also conform to this post-1998 emphasis. The listening section was implemented in 2005.

No written English is required of examinees on this test (although so-called productive skills are required on most second-stage university tests). Given the massive number of candidates who take the Senta Shiken (again, over 500,000 in 2007 [Daily Yomiuri]), the requirement for absolute objectivity, and the demand for swift grading, the absence of productive tasks is not surprising.

Discussion of the 2006 Senta Shiken

The text-types and their weighting immediately reveal the 2006 test's priorities and focus. The reading section shifts gradually from a limited, discrete-item focus to extended, expository texts (word-sentence-paragraph-short essay). Another way of describing this is to say that the test moves from knowledge to skills to comprehension. It is noteworthy that in the reading section far more weight (almost 75%) is placed upon the more extended, comprehensive, integrated texts and tasks than upon discrete items and sentence-level or lower tasks (146 vs. 54 points). This is largely because the later extended-reading sections, with a large number of passage-dependent tasks, often have values of six or seven points per task/question as compared to 2 points for the majority of items in the discrete-item sections 1 and 2. This means that specific grammatical focus is evident in only about one-fifth of the test in total (and even within the section that falls under the rubric of "grammar" we can find a focus upon more communicative aspects of language, such as norms of social interaction and uptake). This makes the *Senta Shiken* decidedly not a "grammar test." The tasks on the 2006 test demand a variety of wide-ranging and comprehensive reading skills: making inferences, summarizing, recognizing themes, extrapolating information indirectly, knowing the functions of rhetorical signals and connectors, and applying sociopragmatic knowledge. In order to complete these varied tasks, comprehensive, holistic reading skills are required, as opposed to mere knowledge of English minutiae.

The topics and genres of the texts in 2006 are varied: journalistic essays on wide-ranging subjects, narratives, dialogues, opinion pieces, and so forth. Several include visual prompts. Having a wide variety of topics and genres increases the likelihood of the test content appealing to a

broader range of examinees' knowledge areas and interests and thereby increases the validity of the test by not catering to limited, specific knowledge or familiarity with certain content. Furthermore, the variety of task types provides a more accurate measure of wide-ranging English skills, with the added benefit that they would appeal to different learner types.

This variety of texts and tasks also applies to the listening section where we find monologues (speeches, narratives, announcements), dialogues (formal and casual, extended or short), and visual prompts, which demand a wide variety of skills such as listening for gist, listening for specific information, making deductions and inferences, and manipulating data-formats that might appeal to a wide variety of learner types.

Analysis of the 1981 Senta Shiken English Exam

A copy of the 1981 exam was obtained by special request from the National Library, Nagatacho, Tokyo. The 1981 test is divided into eight sections, all reading-based. The total Again, value is 200 points.

Section 1: 50 points

- *Text type*: 25 decontextualized sentences
- *Topics/Genre*: none identifiable
- *Task type*: fill-in-the-blank from multiple choices of four (25 questions)
- *Skills required*: discrete grammatical and /or lexical knowledge

Section 2: 15 pts.

- *Text type*: five short exchanges (two or three turns each)
- *Topics/Genre*: general conversation (various)
- *Task type*: fill-in-the-blank at the sentence / phrasal level; one question for each exchange
- *Skills required*: understanding pragmatic force, prediction, interactive and conversational cohesion

Section 3: 11 pts.

- *Text type*: single words
- *Topics/Genre*: none

- *Task type*: spot the different item from a multiple choice selection of four (11 questions)
- *Skills required*: knowledge of pronunciation at a phonemic level

Section 4: 24 pts.

- *Text type*: six decontextualized declarative sentences
- *Topics/Genre*: none identifiable
- *Task type*: sequencing, ordering
- *Skills required*: word order (at a syntactical level)

Section 5: 12 pts.

- *Text type*: one short conversation/ one short essay
- *Topics/Genre*: personal conversation/ education in England (expository article)
- *Task type*: reorder sentences into cohesive paragraphs (one weighty question in each paragraph)
- *Skills required*: organizing discourse, making predictions, understanding rhetorical connections

Section 6: 27 pts.

- *Text type*: three short paragraphs on very different topics in a magazine article style
- *Topics/Genre*: a) rising early (narrative), b) satellites, c) freedom of speech (expository paragraphs)
- *Task type*: a) paraphrasing proverbs (2 questions), b and c) true/ false (one of 5five) summarizing questions
- *Skills required*: proverb knowledge, summarizing

Section 7: 21 pts.

- *Text type*: very short article of about 200 words
- *Topics/Genre*: birdsong (magazine/ journal form)
- *Task type*: paraphrasing; summarizing
- *Skills required*: understanding rhetorical connections, paraphrasing, summarizing

Section 8: 40 pts.

- *Text type:* paragraph of about 250 words
- *Topics/Genre:* expository journal article on reading habits
- *Task type:* a) four true / false questions (one of five paraphrased statements based on the essay is true) b) one two-part question on summarizing the essay theme.
- *Skills required:* paraphrasing, summarizing

Discussion of the 1981 Senta Shiken

In the 1981 exam's extended reading and/or expository sections, the tasks invariably require examinees to paraphrase underlined passages or to look for specific information. This means that rather than comprehending rhetorical flow, cohesion, or the development of themes, one often needs only to read the area that contains the key phrase or word in order to answer the question (passage-independent). Therefore, many questions in 1981 that nominally appear to be integrative and holistic actually have a discrete-item focus.

The extended /expository reading texts in 1981 are also notable in that they seem to reflect the interests and reading habits (actual or idealized) of the test-makers themselves, invariably rather formalized magazine journal-based passages which often appear to be extracted from larger original texts. Having such a narrow content or stylistic focus is not evident on the 2006 exam.

It is also readily apparent that several tasks on the 1981 exam test knowledge of specific proverbs (Q# 6.1), and obtuse expressions. Q# 6.3's key phrase (referring to freedom of speech) reads:

"...the way to its attainment has lain through lakes of blood"

Some require specific *cultural* knowledge:

Q# 1.12 What they call "the first floor" in America is called "the _____ floor" in England.

1. base 2. ground 3. primary 4. second

Some employ rather stilted forms:

Q# 1.9 He knows little of math, _____ of chemistry.

1. as well as 2. still less 3. no less than 4. still more

Some texts lack clear contexts, which effectively invalidates the tasks:

Q# 2.2 *“Why don’t you cut your cake in half?”*

“_____ . If I cut it in half I won’t be able to tell which is the biggest.”

1. *Because it is too big for me*
2. *All right, I’ll cut it in half*
3. *Because I’d like to eat only a half of it*
4. *Because I’d like to eat the biggest piece last*

Many language forms are rather arcane (“*little boy*” being used as an address form), with some exchanges awkwardly contrived (“*I tore them up*”), both as noted in Q# 2.1:

Q# 2.1 *“I suppose you didn’t see some papers on the table in my room, little boy?”*

“_____” . “What! You must be taught once and for all not to touch what does not belong to you.”

- | | |
|---------------------------|-------------------------------------|
| 1. <i>Yes, I did.</i> | 2. <i>No, I didn’t.</i> |
| 3. <i>I tore them up.</i> | 4. <i>I left them as they were.</i> |

The 1981 test also includes numerous poorly designed distractors (such that the correct answer can be deduced without referring to the actual text at all). No contextually barren passages nor such arcane/obtuse language were found on the 2006 test.

Earlier we mentioned that “difficulty” in terms of reading scales would not be a factor in this study. However, if one considers that arcane language forms, decontextualized passages, and unnatural interactive speech (qualities that standard reading difficulty scales often fail to measure) compound difficulty, then the 1981 has “difficulty” in spades.

Comparison of the 1981 and 2006 *Senta Shiken*

Several differences between the two tests become immediately evident. Most salient among these are:

- The lack of a listening component in 1981.
- On the 1981 test, 85 points (42% of the total) are given to questions that demand discrete, decontextualized knowl-

edge or skills at the sentence level or lower. Compare this to 54 points (21.6%) in the 2006 exam.

- The 1981 exam also shows much less variation in terms of text type. Five of the six expository texts in 1981 are of the magazine / journal article format, unlike the extended conversations, dialogues, narratives, visual prompts, and so forth encountered in the 2006 test.
- The 2006 texts are much longer. Whereas the 1981 exam has only one text of 250 words or more, the 2006 exam contains four, three of them nearing 500 words. This allows for greater contextualization and demands more holistic and integrative reading skills.
- While in the 2006 discrete-item sections there is a marked concern with features of coherence, such as logical connectors and discourse markers, the same sections in 1981 focus upon narrower units of discourse, such as prepositions and phrasal verbs.
- Of the 22 questions asked in the essay sections of the 2006 exam only four (17%) can be reasonably said to be answerable without comprehending wider contexts (largely passage independent). Of the 13 questions tied to similar texts in 1981, eight (63%) can be answered by reading only a small section of the text (largely passage dependent).
- The extended / expository texts in 1981 average only 2.4 questions per text (12 questions for five texts). The 2006 test not only averages 3.8 questions per text (15 questions for four main texts) but a greater variety of questions are asked, demanding a greater variety of reading skills—generally more holistic and integrative. This means that there is more “wasted” text in the 1981 test (section eight being the exception).

None of this is to say that the 1981 test was completely bereft of positive qualities, but when compared proportionately to the 2006 exam, the construct validity is palpably lower.

Conclusions

By many standard measures of test validity (utilizing a wide variety of texts, a multidimensional task focus, an emphasis upon higher level cognitive skills, fewer cases of arcane, decontextualized, obtuse and narrow-focus texts and tasks) the 2006 test is clearer the superior test. Moreover, the skills addressed on the test are also indicative of sound reading and listening pedagogy and appeal to a wide range of learner types and learning strategies.

In addition, the 2006 test realizes many of Brown's (2000) recommendations. For instance, a) there is a wide variety of samples and types; b) it tests those abilities that one would like to develop and encourage—that is, reading and listening for higher-level meanings and comprehension, rather than focusing on discrete items which would encourage a narrow item-by-item, bottom-up translation approach; c) there is a great variety of examination formats; d) the test puts a higher emphasis on assessing higher order cognitive skills; and e) the test's texts and tasks are not at all limited to academic fields.

If there is a washback effect from the *Senta Shiken* it should be a positive one. In order to enable them to succeed on the exam, students should be made to practice and develop a number of differing skills such as predicting, summarizing, expounding, extrapolating, interpreting, sequencing, reconstructing, paraphrasing, reading and listening for specific information, and reading and listening between the lines, all using texts of a variety of genres (narratives, information transactions, casual dialogues, scientific essays, personal essays, etc.). The listening section also demands a variety of developmental listening strategies since tasks and the skills required vary from making inferences to listening for gist and/or listening for specific information.

The 1981 *Senta Shiken* does conform to the critical view of a narrow, grammar-based, discrete-item test which does not adequately address more progressive or holistic educational strategies. However, this stands in sharp contrast to our findings regarding the 2006 exam. The various revisions and changes in MEXT's policies and practice regarding entrance exams do seem to be bearing some fruit and are not inconsistent with its broader educational aims nor with sound pedagogical and testing theory.

Michael (Mike) Guest is Associate Professor of English in the Medical School of Miyazaki University. He also serves as a section editor for the ETJ (English Teachers in Japan) Journal and is a regular columnist on EFL teaching for the national Daily Yomiuri newspaper. His interests include lexis, culture, and critical thinking, as well as testing.

References

- Arai, K. (1999). Articulation to shite no daigaku nyushi [University entrance examination as a way of articulation]. *IDE–Gendai no Koutou Kyouiku [IDE–Higher Education in Modern Society]*, 408, 13–19.
- Bailey, K. M. (1999). *Washback in language testing* (TOEFL Monograph Series No. 15). Princeton, NJ: Educational Testing Service
- Brown, J. D. (2000). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken: JALT Testing & Evaluation SIG newsletter*, 3 (2), 4-8.
- Brown, J. D. (2002). English language entrance examinations: A progress report. *Proceedings of the 1st Annual JALT Pan-SIG Conference*, (pp. 94-104.).
- Brown, J. D., & Yamashita, S. (1995). English language entrance exams at Japanese universities: What do we know about them? *JALT Journal*, 17 (1), 7-30.
- The Daily Yomiuri*. (2007, Jan. 20). p. 2
- Gorsuch, G. (1998). Yakudoku EFL instruction in two Japanese high school classrooms: An exploratory study. *JALT Journal*, 20 (1), 6-32.
- Guest, M. (2000). ...““But I have to teach grammar!”: An analysis of the role “grammar” plays in Japanese university English entrance examinations. *The Language Teacher*, 24 (11), 23-29.
- Guest, M. (2006). Teaching progressively...for the *Senta Shiken!* In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), *JALT 2005 Conference Proceedings*, 1182-1191 [CD-ROM]. Tokyo: JALT.
- Guest, M. (2007). A Response to Yoko Ichige’s “Validity of Center Examinations for Assessment of Communicative Ability”. *On Cue*, 15 (1), 28-32.
- Hirezaki, K. (1991). *Daigaku nyuushi–Kaikaku no kokoromi [University entrance examination: Attempts at its improvement]*. Tokyo: Iwanami Booklet.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge University Press.
- Ichige, Y. (2006). Validity of center examinations for assessment of communicative ability. *On Cue*, 14 (2), 13-22.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 28 (1), 77-96.
- Kuroha, R. (1992). 1960 Nendai ikou no daigaku seisaku–Sono taiken teki seiri to kentou [Policy of higher education after the 1960s: Its pigeonholing and review through personal experience]. *Daigaku Kenkyu [University Studies]*, 81-92.

- McVeigh, B. (2001). Higher education, apathy and post-meritocracy. *The Language Teacher*, 25 (10), 29-32.
- Monbukagakusho (2000). *Daigaku nyuushi no kaizen ni tsuite* [How to improve the university entrance exams]. Retrieved August 21, 2007, from <http://www.mext.go.jp/b?Qmenu/shingi/12/daigaku/toushin/001102.html>
- Monbukagakusho (2003). *The course of study for foreign languages*. Retrieved August 21, 2007, from <http://www.mext.go.jp/english/shotou/index.htm>
- Mori, R. (2002). Entrance examinations and remedial education in Japanese higher education. *Higher Education*, 43, 27-42.
- Mulvey, B. (2001). The role and influence of Japan's university entrance exams: A reassessment. *The Language Teacher*, 25 (7), 11-15.
- Stout, M. (2003). Not guilty as charged: Do the university entrance exams in Japan affect what is taught? *ETJ Journal*, 4 (1), 1-7.
- Trelfa, D. (1998). The development and implementation of education standards in Japan. In *The Educational System in Japan: Case Study Findings*. Retrieved August 21, 2007, from <http://www.ed.gov/pubs/JapanCaseStudy/chapter2d.html>
- Watanabe, Y. (1996). Does grammar translation come from the entrance exams? Preliminary findings from classroom-based research. *Language Testing*, 13 (318-333).