

Articles

English Language Entrance Examinations at Japanese Universities: What Do We Know About Them?

James Dean Brown

University of Hawaii at Manoa

Sayoko Okada Yamashita

International Christian University

Many types of items are used in language tests, ranging from receptive to productive and discrete-point to integrative, and passage-based to passage-independent. This study defines and investigates the types of items used in entrance examinations in English administered at Japanese universities in 1993. Ten examinations each from private and public universities along with the nationwide "Center" examination (21 examinations) were studied to determine: 1) the difficulty of the various reading passages; 2) the differences in the levels of reading passage difficulty in private and public examinations; 3) the types of items, their variety, and the test length; 4) the differences in the types of items used in private and public examinations, and 5) the skills measured. The examinations were analyzed item-by-item by a native speaker of English and a native speaker of Japanese. Computer analysis was performed to investigate the level of difficulty of the reading passages. The results should help English teachers in Japan prepare students for taking such tests and help their students in deciding on which to take. Equally important, this study may aid those responsible for creating entrance examinations to prepare high quality tests.

言語能力のテストでは様々な項目が使われている。理解から産出、細かい項目を見るものから統合的なもの、文章を使ったものを使わないものなどその幅は広い。この研究は、英語の1993年に日本で行われた大学入試で使われた問題のタイプを分類・検討する。分析の対象としたのは、私立大学、国公立大学の入試問題を各10校分および大学入試センター試験である。分析項目は、1) 読解文の難易度、2) 私立大学と国公立大学における読解文の難易度の違い、3) 問題のタイプ、多様性、試験の長さ、4) 私立大学と国公立大学におけ

る問題のタイプの違い、5) 測定されているスキルである。問題は1題ずつ英語母語話者と日本語母語話者によって分析され、読解文の難易度はコンピュータ処理をした。この結果は、日本で教える英語教師が、学生がこれらの試験を受けるための準備をし、どのテストを受けたらいいかを助けるために役にたつであろう。さらに、入試問題の作成者にとっては、質の高いテストを作るための助けとなるかもしれない。

In Japan, prestigious secondary schools and most universities require aspiring students to take entrance examinations created by the institutions themselves. The result is a phenomenon which is known as *shiken jigoku*, or examination hell, which describes the months and years that Japanese young people spend preparing for entrance examinations.

Is examination hell a new phenomenon? Amano (1990, p. xx) says that it existed as early as the 1920s. His book, which details the history of examinations in modern Japan, traces the roots of the present examination system back to the Meiji Restoration and the process of modernization. As Cummings points out: "Over the postwar period the number of children seriously committing themselves to exam preparation has steadily increased, whereas the number of openings in elite schools has scarcely changed" (1980, p. 206). In fact, over the years, the increasing competitiveness has been accompanied by increases in the number of universities from 45 in 1945 to 201 in 1950 (Amano, 1990, p. xv) followed by another nearly fivefold rise in the next 45 years: "Every year about 45 percent of Japan's high school graduates seek enrollment at 996 post secondary institutions: 460 four-year colleges and 536 two-year colleges" (Shimahara, 1991, p. 126).

Regardless of the numbers of universities involved, "examination hell" exists in Japan, and does so in a big way. Transitions between junior and senior high school as well as between high school and university are marked by entrance examinations. Most Japanese believe that their success and the success of their children hinge on passing these examinations. As a result, ". . . families devote a surprising proportion of their resources toward assisting their children in exam preparation, and children devote long hours day after day to study" (Cummings, 1980, p. 206).

This is not to say that the Japanese are necessarily happy with the system. Frost's (1991) "Examination Hell" provides an excellent overview of the reasons why the Japanese criticize and bemoan, yet feel they need, and therefore admire and put up with it. In brief, the exami-

nation system is widely criticized for having “undesirable effects on curriculum, on foreign language instruction, on family life, and on children’s emotional, physical, and intellectual development” (Tsukada, 1991, p. 178).

Unfortunately, as pointed out by Frost, “separate university achievement tests simply have too long a history and meet too many needs ... to disappear simply because a new generation is beginning to be truly worried about them” (1991, p. 303). As a result, the system is probably here to stay—at least for the time being. Since this system dramatically affects English teaching in Japan, it is important for EFL teachers to know as much as possible about these examinations—particularly the English examinations.

Definitions

Before considering these examinations in detail, we must define some key terms: What is a test item? What is the difference between discrete-point test items and integrative ones? What is the distinction between receptive test items and productive ones? And how do translation items relate to all of the above item types? Answers to these questions have a direct bearing on any study that purports to investigate what a set of tests is doing.

What is an item? A test item is defined in this study as the smallest distinctive unit on a test that yields separate information; that is, the smallest part that could be scored separately (after Brown, in press). Thus, a single multiple-choice item is considered a test item, but so is an essay, or a translation task, because each is the smallest part that can yield a score.

Discrete-point versus integrative. A discrete-point item is one that is designed to test a single well-defined language point. Examples include true-false, multiple-choice and matching; fill-in items are often discrete-point in nature. In the early days of language testing (and even today in some quarters), the sum of the discrete-point items on a test was thought to represent overall language proficiency (Spolsky, 1978; Hinofotis, 1981). Carroll (1972) threw this simplistic early belief into question by proposing that language might be integrative.

By definition, if language is integrative, the various components of language (phonological, lexical, syntactic, semantic, pragmatic, etc.) interact in complex and sometimes unpredictable ways. To accommodate this notion, a new class of language tests began to be used: integrative tests. The items on an integrative test are sometimes harder to define

and identify because, being situated in a language context, they may interact with each other and with other elements of the language context in relatively complex ways. Examples of integrative tests are dictations, cloze tests, essay writing tasks, and interviews. Such tests require quite different skills from those required in taking a discrete-point test.

Generally, this discrete-point versus integrative distinction is not taken to be an absolute dichotomy. Rather, any given item type will fall on a continuum somewhere between completely discrete-point on one end and integrative on the other. Nonetheless, the distinction has helped us understand what we are doing in language testing and may help in investigating what is being tested.

Receptive versus productive. Receptive item types are those in which one receives language in the sense of reading it or hearing it, but is not required to produce language (i.e., to speak or to write anything). Generally, true-false, multiple-choice, and matching items are the types that would be considered receptive under this definition. The examinees read or hear a bit of language and are then given a choice of some sort from among options that they hear or read, and make a mark or circle something on an answer sheet. Clearly, in answering receptive item types, one need not produce anything at all in the language. Such items are probably best suited for testing the receptive language skills of reading and listening, and perhaps language knowledge like grammar and vocabulary.

Productive item types are different from receptive items in that they that language actually be produced in some form or other. While productive items usually include some receptive language (in the form of directions or prompts), they all require the production of written or spoken language. Generally, fill-in, short answer, and task-oriented (e.g., compositions, interviews, role plays) items are productive items because, though they do have elements of receptive language, ultimately one must produce some written or spoken language in order to answer them. Such items are probably best suited for testing the productive skills of speaking and writing, as well as the communicative interactions of reading with writing, or listening with speaking. Presumably, given the current emphasis on communicative language teaching, at least some productive item types should be used in almost any language test. By using them we send our students a signal that we send to the students is that we care about their abilities to use not only the receptive skills of reading and listening and those of language knowledge, like grammar and vocabulary, but also the productive skills of writing and speaking, i.e., their actual abilities to use language to communicate ideas, needs, and wants to other people.

Translation items. Translation items require examinees to translate a phrase, sentence, or paragraph from their first language (L1) into their second language (L2), or vice versa. For instance, a student might be required to write out a Japanese translation for a sentence given in English.

In English speaking countries and the other countries of the European Community, translation is generally considered a very special skill. Translation of documents and simultaneous interpretation are skills that are taught to those who have achieved a very high level of proficiency in both their first and second language. Such training requires years of study to complete. In this view, effective translation is a task that most advanced speakers of an L2 may not be qualified to do, much less junior or senior high school students. More importantly from a pedagogical perspective, seeing translation tasks on a test allows our students to infer that translation is a legitimate strategy to use in trying to communicate. Even if they are defensible from a pedagogical perspective, translation items are not practical in the many ESL settings where multiple L1s are represented in every classroom. For all of the above reasons, it is our experience that translation items appear to have virtually disappeared in ESL settings—except of course in the very specialized graduate schools that train translators.

Purpose

As the purpose of this study was to investigate the Japanese university English entrance examinations, our goals were: 1) to describe current testing practices at major institutions and 2) to establish a baseline of information so that change or lack of change in the testing practices of such universities can be monitored in future years. To these ends, the following more formal research questions were posed:

1. How difficult are the various reading passages?
2. Are there differences in the levels of reading passage difficulty on private and public university examinations?
3. What types of items are used, how varied are they, and how does test length vary?
4. Are there differences in the types of items and test lengths found in private and public university examinations?
5. What skills are measured on these examinations?

We hope answers to these questions will provide information to English teachers in Japan that will help enable them to prepare students

for such tests. Equally important, this study should encourage those responsible for creating the English entrance examinations at various universities to produce practical, reliable, and valid tests.

Method

Materials: In making our selections we wanted 10 each of the most prestigious private and public universities, as well as a good geographical spread. Our two primary sources of information were two guidebooks, published annually and readily available to the general public. Each contained a number of English entrance examinations given in 1993 along with helpful hints in Japanese for preparing for the examinations. Koukou-Eigo Kenkyuu, 1993a included 92 examinations from as many private universities throughout Japan, from which 10 were selected for this study: Aoyama Gakuin University, Doshisha University, Keio University, Kansai Gaidai (Foreign Languages) University (KANGAI), Kansai University, Kyoto University of Foreign Studies (KYOUPS), Rikkyo University, Sophia University, Tsuda University, and Waseda University. Koukou-Eigo Kenkyuu, 1993b provided a collection of 64 examinations from public universities, 49 of which were national and 15 municipal, and one was the nationwide "Center" examination. From this book, 10 public universities were selected. Eight of these were national: Hitotsubashi University (HITOTSU), Hokkaido University, Kyoto University, Kyushu University, Nagoya University, Osaka University, University of Tokyo, and Tokyo University of Foreign Studies (TYOUPS). Two universities were municipal ones: Tokyo Metropolitan University (TORITSU) and Yokohama City University.

The *Daigaku Nyuushi Sentaa* (University Entrance Examination "Center") examination was also selected. (The "Center" examination is a first-stage exam, somewhat analogous to the College Board SAT, in that many universities subscribe to it.) It was included because it is administered nationwide and is used by a variety of universities, solely or in combination with their own, for a total of 21 tests. (The "Center" examination is a first-stage exam, somewhat analogous to the College Board SAT, in that many universities subscribe to it.)

Procedures: Each examination selected was studied for various characteristics. Item-by-item, we labeled the type of item, the purpose of the item, the number of options given to students, the language(s) involved in the item, and the task involved.

Data entry took several forms. In all cases, however, the computer programs involved were for IBM (MS-DOS) computers. First, each of

the items was coded for item type and recorded in the *QuattroPro™* spreadsheet program (Borland, 1991). Second, all of the reading passages were typed into the *WordPerfect™* (WordPerfect, 1988) computer program.¹

Analyses: All of the English language reading passages (without the directions in Japanese) on the entrance examinations were analyzed using the *RightWriter™* computer program (Que Software, 1990) for such features as number of words, number of unique words (i.e., number of words, counting each word only once), percentage of unique words (type-token ratio), number of sentences, syllables per word, and words per sentence (and many other characteristics beyond the scope of this paper). In addition, the Flesch, Flesch-Kincaid, and Fog readability indexes were calculated. The number of words, number of sentences, syllables per word, and words per sentence are self-explanatory. However, the other statistics may not be so clear. The number of unique words is the number of different words used in a passage, counting each word only once regardless of the number of times it appears. The Flesch, Flesch-Kincaid, and Fog readability indexes are measures which estimate the difficulty of the reading level of the passage. The Flesch scale ranges from 0 to 100; the higher the number, the easier the passage should be to read. The Flesch-Kincaid and Fog readability indexes are both expressed as the grade level of students for which the reading passages should be appropriate. In other words, a Flesch-Kincaid or Fog index of 10.5 for a given passage would indicate that the passage would be appropriate for an average student about halfway through the tenth grade in an American school system. (For more on these readability indexes, including the formulas, see Que Software, 1990; Flesch, 1974.)

The analysis of item types was done with the spreadsheet program. Simple descriptive statistics were used to make comparisons between universities and types of universities. These statistics consisted primarily of averages and percentages. Because of the descriptive nature of this study, no inferential statistics were deemed necessary.

Results

1. How difficult are the various reading passages used in these university English language entrance examinations?

Since all of the examinations used reading passages as the basis for at least some of the items on the test, we began by looking at those passages.

Tables 1A (private) and 1B (public) indicate that all of the universities used at least one passage and that two of the public universities, TOKYO and TYOUPS, used as many as six passages. The average lengths of the passages were ascertained by examining the number of words. While TOKYO and TYOUPS had several passages, they were relatively short, 220

Table 1a: Reading Passage Statistics for Private University Entrance Examinations, 1993 (Averages)

Statistic	AOYAMA	DOSHISHA	KEIO	KANGAI	KANSAI	KYOTOUPS	RIKKYO	SOPHIA	TSUDA	WASEDA
No. of Passages	2	2	1	2	3	1	2	4	2	4
Words	381.50	588.00	986.00	385.00	488.00	863.00	388.00	431.00	453.00	438.00
Unique Words	211.50	296.50	515.00	200.00	239.00	401.00	194.00	237.25	212.00	219.50
Type-Token Ratio	55.44	50.43	52.23	51.95	48.98	46.47	50.00	55.05	46.80	50.11
Syllables/Word	1.49	1.48	1.65	1.50	1.46	1.53	1.47	1.54	1.46	1.53
Sentences	21.00	26.50	51.00	25.50	28.67	48.00	20.00	20.25	33.50	26.50
Words/Sentence	17.57	22.34	19.04	15.53	18.98	17.75	19.42	21.81	21.31	16.50
Flesch	63.59	60.61	48.08	64.22	65.47	59.49	64.29	55.86	61.77	60.60
Flesch-Kincaid	8.59	9.94	11.28	8.16	8.40	9.28	8.83	10.51	9.92	8.87
Fog	9.98	11.84	13.26	10.07	9.92	10.71	11.34	12.56	11.51	10.64

Table 1b: Reading Passage Statistics for Public University Entrance Examinations, 1993 (Averages)

Statistic	HITOTSU	HOKKAIDO	KYOTO	KYUSHU	NAGOYA	OSAKA	TOKYO	TORITSU	TYOUPS	YOKOHAMA
No. of Passages	2	3	2	3	3	4	6	2	6	3
Words	621.50	479.67	417.50	451.33	279.33	289.50	220.00	474.00	264.50	286.00
Unique Words	295.50	244.00	227.00	212.00	166.00	145.50	125.17	251.50	143.50	159.33
Type-Token Ratio	47.55	50.87	54.37	46.97	59.43	50.26	56.89	53.06	54.25	55.71
Syllables/Word	1.43	1.53	1.62	1.42	1.60	1.45	1.50	1.52	1.48	1.64
Sentences	34.50	35.00	18.50	33.67	11.67	15.50	11.17	23.00	16.17	12.67
Words/Sentence	16.47	14.91	22.38	18.56	20.33	20.26	25.87	20.64	16.65	25.71
Flesch	69.23	62.51	47.60	68.30	50.44	64.08	54.28	58.52	65.48	42.47
Flesch-Kincaid	7.70	8.24	12.18	8.34	11.28	9.27	11.92	10.01	8.03	13.61
Fog	10.18	9.84	14.54	10.80	12.75	11.75	14.29	12.40	9.45	15.91

and 264.5 words per passage on average, respectively. In contrast, two of the private universities, KEIO and KYOTOFU, used only one passage each, but used relatively long ones of 986 and 863 words, respectively.

The number of unique words, or number of different words used, is an indication of the variety of vocabulary items presented. The percentage of unique words, or type-token ratio, was calculated by dividing the number of different words by the total number of words. This percentage can therefore be considered a clearer indication of the variety of vocabulary because, as a proportion, it is therefore more easily comparable across passages, universities, and university categories. The number of syllables per word is a rough indication of the difficulty of the words (based on the notion that longer words are generally more difficult than shorter ones).

The number of sentences is fairly straightforward to interpret. For instance, if two passages are approximately the same length in words but one has more sentences than the other, the first passage must have shorter sentences. The issue of sentence length is actually easier to address by examining the number of words per sentence because that statistic describes the average length of the sentences. However examined, sentence length can be considered a rough indication of the complexity of the syntax involved in the passages.

The Flesch readability index indicates that the passages ranged in overall reading level from "standard" (69.23) at HITOTSU to "difficult" (42.47) at YOKOHAMA. The Flesch-Kincaid and Fog readability indexes indicate that the passages would be appropriate for native speakers ranging generally from eighth grade (about 13 years old) at TYOVS to late fifteenth grade level (third year of university, or about 21 years old) at YOKOHAMA (15.91).

2. Are there differences in the levels of reading passage difficulty in private and public university examinations?

Tables 1A and 1B show numerous differences among the universities within the private and public categories as well as between those two categories. There is no space to consider all of these differences here. However, the summary statistics of averages shown in Table 2 reveal some interesting overall contrasts between the private and public universities. The results for the Center examination are also shown in Table 2 so that they can be directly compared to the average private and public examinations and to the averages for all 21 examinations (TOTAL).

The public universities set more reading passages, but shorter ones, than the private universities, while the Center examination was some-

Table 2: Reading Passage Statistics Summarized by University Type (Averages)

Statistic	Private	Public	Center	Total
No. of Universities	10	10	1	21
No. of Passages	2.30	3.40	3	2.86
Words	540.15	378.33	178.33	445.87
Unique Words	272.58	196.95	100.67	228.38
Type-Token Ratio	50.74	52.94	53.88	51.84
Syllables/Word	1.51	1.52	1.41	1.51
Sentences	30.09	21.18	9.33	24.86
Words/Sentence	19.03	20.18	17.01	19.48
Flesch	60.40	58.29	70.35	59.87
Flesch-Kincaid	9.38	10.06	7.67	9.62
Fog	11.18	12.19	10.17	11.61

where in between in terms of number but has considerably shorter passages. It also seems that, in terms of percentage of unique words (type-token ration), syllables per word, and words per sentence, there were no particularly interesting differences between the private, public, and Center examinations. However, the average readability levels of the passages vary by type of university. On the Flesch scale, the average private university passage was "standard" (60.40) while the average public university was "fairly difficult" (58.29) and the Center examination was "fairly easy." Using the Flesch-Kincaid scale, the private-university passages were of about 9th grade reading difficulty, the public universities about 10th grade and the Center examination average late 7th grade. The same pattern of relative difficulty occurred in the Fog scale results. In general then, students should find the passages easiest on the Center examination, a bit more difficult on the private university examinations, and most difficult on the public university examinations.

3. What types of items are used, how varied are they, and how does test length vary?

Item types. Tables 3A, private universities, and 3B, public, summarize the different types of items. The percentages are provided so that easier comparisons can be made between universities without the confusion caused by differing test lengths.

Within the frequencies and percentages, the item types are classified by skill area with reading and writing collapsed together to represent written tests; translation is treated as a separate skill (as discussed above), and listening stands alone. Within each skill area, we found somewhat

different types of items. Reading/writing skills contained multiple-choice, rephrasing or reordering, fill-in the blank, and short answer or essay items, with no true-false or matching items. Within translation, we found English to Japanese translation and, somewhat less frequently, Japanese to English. Among the listening items we found true-false, multiple-choice, fill-in, and dictation items.

Item variety. Tables 3A and 3B provide a great deal of information. We cannot begin to interpret all of it here in prose. However, certain facts do jump out of these tables. Notice, for instance, that different universities use different combinations of item types.

DOSHISHA, KANGAI, KYOTOUFS, and SOPHIA placed great emphasis on multiple-choice items. KEIO and TORITSU placed more emphasis on translation (nearly 53 percent), and HITOTSU and KYOTO emphasized multiple-choice and fill-in listening items. The one thing that seems clear, at this point, is that the character of the various examinations in terms of item types can vary tremendously. Thus the types of items that candidates will face depends to a great degree on which examination(s) they choose to take.

However, as with all classifying and categorizing, the process of analyzing the item types oversimplifies reality. Within each of these apparently simple categories there was additional variation. Consider just the first category of reading/writing multiple-choice items. There was tremendous variation in this seemingly straightforward item type. The number of options supplied ranged from two to six. Some of the options were in English, some in Japanese. Some of the items were based on the reading passages and some stood alone. Some items posed a straightforward question; others required one to select the option that successfully filled a blank. Still others required the selection of the option that matched a word or phrase, and finally, some provided context for blanks in the form of multiple-choice cloze. All in all, with the various combinations of the factors just discussed, we found 16 different categories of passage-based multiple-choice items and six categories of multiple-choice items that stood alone for a total of 22 different types of multiple-choice items. However, even this is an oversimplification. Similar variety existed within each of the other types of items shown in Tables 3A and 3B. We are not the first to have noticed this phenomenon. As Duke (1986) put it, ". . . written English tests in [Japanese] public schools can be quite intricate. They often require detailed knowledge of the techniques of the language, both written and oral" (p. 154).

One ramification of this intense variation of item types on most of the examinations is that the students are forced to change item types often within any given test. This means that directions (always in Japa-

Table 3a: The Variety of Item Types on Private University Examinations, 1993

SKILL: ITEM		TYPE	Aoyama	Doshisha	Keio	Kangai	Kansai	Kyotoufu	Rikkyo	Sophia	Tsuda	Waseda	
<i>Frequencies</i>	Reading/Writing:	Multiple-Choice	10	56	1	69	32	41	15	61	17	19	
		Rephrase/Reorder	0	0	5	0	2	4	3	0	0		
		Fill-In	0	0	2	0	15	0	20	0	7	13	
		Short-Answer/Essay	1	0	1	0	0	0	1	0	0	0	
	Translation:	Translate (E->J)	4	1	3	0	2	1	1	0	2	0	
		Translate (J->E)	0	2	1	0	1	1	1	0	1	0	
	Listening:	True-False	10	0	0	0	0	0	0	0	0	0	0
		Multiple-Choice	10	0	0	0	0	0	0	0	0	0	0
		Fill-In	0	0	0	0	0	0	0	0	0	0	0
		Dictation	0	0	0	0	0	0	0	0	0	1	0
	Total No. of Items			35	59	8	74	50	45	42	64	28	32
	Time Allowed (min.)			100	100	120	90	90	80	75	90	100	90
	<i>Percentages</i>	Reading/Writing:	Multiple-Choice	28.57	94.92	12.50	93.24	64.00	91.12	35.71	95.31	60.71	59.37
Rephrase/Reorder			0.00	0.00	0.00	6.76	0.00	4.44	9.52	4.69	0.00	0.00	
Fill-In			0.00	0.00	25.00	0.00	30.00	0.00	47.62	0.00	25.00	40.63	
Short-Answer/Essay			2.86	0.00	12.50	0.00	0.00	0.00	2.38	0.00	0.00	0.00	
Translation:		Translate (E->J)	11.43	1.69	37.50	0.00	4.00	2.22	2.38	0.00	7.14	0.00	
		Translate (J->E)	0.00	3.39	12.50	0.00	2.00	2.22	2.38	0.00	3.57	0.00	
Listening:		True-False	28.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		Multiple-Choice	28.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		Fill-In	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		Dictation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.57	
Total Percent			100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	

Table 3b: The Variety of Item Types on Public University Examinations, 1993

SKILL:		ITEM TYPE	HITOTSU	HOKKAIDO	KYOTO	KYUSHU	NAGOYA	OSAKA	TOKYO	TORITSU	TYOUPS	YOKOHAMA	
<i>Frequencies</i>	Reading/Writing:	Multiple-Choice	3	1	0	5	9	4	7	0	6	1	
		Rephrase/Reorder	0	0	0	0	0	1	8	0	6	0	
		Fill-In	0	9	0	8	0	8	6	0	0	0	
		Short-Answer/Essay	2	4	2	6	6	4	2	9	11	7	
	Translation:	Translate (E->J)	7	6	3	4	5	2	3	10	0	2	
		Translate (J->E)	0	4	2	3	3	1	2	0	0	2	
	Listening:	True-False	0	0	0	0	0	0	0	0	0	0	
		Multiple-Choice	10	0	6	0	0	0	0	0	0	0	
		Fill-In	15	0	10	0	0	0	0	0	15	0	
		Dictation	0	0	0	0	0	0	0	0	0	0	
	Total No. of Items			37	24	23	26	23	20	28	19	38	12
	Time Allowed (min.)			120	90	120	120	90	105	120	120	150	90
	<i>Percentages</i>	Reading/Writing:	Multiple-Choice	8.11	4.17	0.00	19.23	39.13	20.00	25.00	0.00	15.79	8.33
Rephrase/Reorder			0.00	0.00	0.00	0.00	0.00	5.00	28.57	0.00	15.79	0.00	
Fill-In			0.00	37.50	0.00	30.77	0.00	40.00	21.43	0.00	0.00	0.00	
Short-Answer/Essay			5.41	16.67	8.70	23.08	26.09	20.00	7.14	47.37	28.95	58.33	
Translation:		Translation (E->J)	18.92	25.00	13.04	15.38	21.74	10.00	10.71	52.63	0.00	16.67	
		Translation (J->E)	0.00	16.67	8.70	11.54	13.04	5.00	7.14	0.00	0.00	16.67	
Listening:		True-False	0.00	0.00	0.00	0.00	0.00	0.00	0.00*	0.00	0.00	0.00	
		Multiple-Choice	27.03	0.00	26.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
		Fill-In	40.54	0.00	43.48	0.00	0.00	0.00	0.00	0.00	39.47	0.00	
		Dictation	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Total Percent			100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00		

* Note that the University of Tokyo had a listening comprehension section on its test, but the items were not given in our source.

nese) are common and prevalent. There can be no question that these examinations assess the students' testwiseness, in that they measure, at least to some degree, the students' abilities to handle varied and novel item types, read directions (in Japanese), and switch gears often. Examination of language tests in the United States, like the TOEFL, CELT, and Michigan Tests, will reveal that they have avoided such issues by keeping the item types similar within fairly large subtests. This practice is based on the belief that tests should assess the students' abilities in the content area (in this case, English) rather than their abilities to take tests.

Test length. As Tables 3A and 3B show, these examinations varied considerably in terms of sheer length, that is, in the time allowed and the numbers of items involved. The amount of time allowed ranged from 75 minutes at RIKKYO to 150 minutes at TYOUFS. The numbers of items varied from a low of eight for KEIO to a high of 74 for KANGAI. Of course, such differences in numbers of items are related in part to the item types involved. For instance, half of the eight KEIO items were translation items (three English to Japanese of 16, 11, and 15 words respectively, and one Japanese to English with 45 characters) which presumably were more demanding than multiple-choice items of the sort which dominated at KANGAI.

Nonetheless, test length, in terms of numbers of items, can be an important factor in determining reliability. Hence, examinations of this importance in the West, like the TOEFL examination, are typically much longer in terms of both numbers of items and time allotted.

4. Are there differences in the types of items and test lengths found in private and public university examinations?

Table 4 summarizes the same information presented in the previous section for the private and public universities grouped together along with the Center examination statistics and the totals for all 21 universities taken together. Hence, the figures presented in this table are averages (except for the Center examination).

On the whole, the same types of items were used at least sometimes in both the private and public universities (with the notable exception of listening true-false and dictation, which were not used at all in the public universities). The Center examination used over 90% multiple-choice items with a few rephrase/reorder types of items providing the only variety.

However, the average balance maintained among the various item types differed widely between the private and public universities. The private universities appeared to rely more heavily on multiple-choice

Table 4: Item Type Variety Summarized by University Type*

Skill:	Item Type	Private	Public	Center	Total
<i>Frequencies</i>					
Reading/Writing:					
	Multiple-Choice	32.10	3.60	48	19.29
	Rephrase/Reorder	1.40	1.50	5	1.62
	Fill-In	5.70	3.10	0	4.19
	Short-Answer/Essay	0.30	5.30	0	2.67
Translation:					
	Translation (E->J)	1.40	4.20	0	2.67
	Translation (J->E)	0.70	1.70	0	1.14
Listening:					
	True-False	1.00	0.00	0	0.48
	Multiple-Choice	1.00	1.60	0	1.24
	Fill-In	0.00	4.00	0	1.90
	Dictation	0.10	0.00	0	0.05
	Total No. of Items	43.70	25.00	53	35.24
	Time Allowed	93.50	112.50	80	103.00
<i>Percentages</i>					
Reading/Writing:					
	Multiple-Choice	63.55	13.98	90.57	41.23
	Rephrase/Reorder	2.54	4.94	9.43	4.01
	Fill-In	16.82	12.97	0.00	14.19
	Short-Answer/Essay	1.77	24.17	0.00	12.36
Translation:					
	Translation (E->J)	6.64	18.41	0.00	11.93
	Translation (J->E)	2.61	7.88	0.00	4.99
Listening:					
	True-False	2.86	0.00	0.00	1.36
	Multiple-Choice	2.86	5.31	0.00	3.89
	Fill-In	0.00	12.35	0.00	5.88
	Dictation	0.36	0.00	0.00	0.17
	Total Percent	100.00	100.00	100.00	100.00

*All statistics for private and public universities as well as total are averages.

items (63.55 percent), while the public preferred short-answer/essay (24.17 percent) and translation (26.29 percent) types. In addition, the private schools averaged 43.7 items per examination, against 25 for the public. The time allotted is also different, with private universities allowing an average of 93.5 and public institutions averaging 112.5.

Table 5a: Categories of Item Types on Private University Examinations, 1993

	ITEM CATEGORY	Aoyama	Doshisha	Keio	Kangai	Kansai	Kyotoufu	Rikkyo	Sophia	Tsuda	Waseda
<i>Frequencies</i>	Discrete-Point	30	56	3	69	47	41	35	61	24	32
	Integrative	1	0	1	5	0	2	5	3	1	0
	Translation	4	3	4	0	3	2	2	0	3	0
	Total No. of Items	35	59	8	74	50	45	42	64	28	32
	Receptive	30	56	3	74	47	43	39	64	24	32
	Productive	1	0	1	0	0	0	1	0	1	0
	Translation	4	3	4	0	3	2	2	0	3	0
	Total No. of Items	35	59	8	74	50	45	42	64	28	32
	Passage-Dependent	12	0	7	14	34	12	17	27	12	18
	Passage-Independent	23	59	1	60	16	33	25	37	16	14
	Total No. of Items	35	59	8	74	50	45	42	64	28	32
	<i>Percentages</i>	Discrete-Point	85.71	94.92	37.50	93.24	94.00	91.11	83.33	95.31	85.71
Integrative		2.86	0.00	12.50	6.76	0.00	4.44	11.90	4.69	3.57	0.00
Translation		11.43	5.08	50.00	0.00	6.00	4.44	4.76	0.00	10.71	0.00
Total % of Items		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Receptive		85.71	94.92	37.50	100.00	94.00	95.56	92.86	100.00	85.71	100.00
Productive		2.86	0.00	12.50	0.00	0.00	0.00	2.38	0.00	3.57	0.00
Translation		11.43	5.08	50.00	0.00	6.00	4.44	4.76	0.00	10.71	0.00
Total % of Items		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Passage-Dependent		34.29	0.00	87.50	18.92	68.00	26.67	40.48	42.19	42.86	56.25
Passage-Independent		65.71	100.00	12.50	81.08	32.00	73.33	59.52	57.81	57.14	43.75
Total % of Items		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 5b: Categories of Item Types on Public University Examinations, 1993

	ITEM CATEGORY	HITOTSU	HOKKAIDO	KYOTO	KYUSHU	NAGOYA	OSAKA	TOKYO	TORITSU	TYOUFS	YOKOHAMA
<i>Frequencies</i>	Discrete-Point	28	10	16	13	9	12	13	0	21	1
	Integrative	2	4	2	6	6	5	10	9	17	7
	Translation	7	10	5	7	8	3	5	10	0	4
	Total No. of Items	37	24	23	26	23	20	28	19	38	12
	Receptive	13	10	6	13	9	13	21	0	12	1
	Productive	17	4	12	6	6	4	2	9	26	7
	Translation	7	10	5	7	8	3	5	10	0	4
	Total No. of Items	37	24	23	26	23	20	28	19	38	12
	Passage-Dependent	12	12	5	11	20	9	13	14	17	10
	Passage-Independent	25	12	18	15	3	11	15	5	21	2
	Total No. of Items	37	24	23	26	23	20	28	19	38	12
	<i>Percentages</i>	Discrete-Point	75.68	41.67	69.57	50.00	39.13	60.00	46.43	0.00	55.26
Integrative		5.40	16.67	8.70	23.08	26.09	25.00	35.71	47.37	44.74	58.33
Translation		18.92	41.66	21.73	26.92	34.78	15.00	17.86	52.63	0.00	33.34
Total % of Items		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Receptive		35.14	41.67	26.09	50.00	39.13	65.00	75.00	0.00	31.58	8.33
Productive		45.94	16.67	52.17	23.08	26.09	20.00	7.14	47.37	68.42	58.33
Translation		18.92	41.66	21.74	26.92	34.78	15.00	17.86	52.63	0.00	33.34
Total % of Items		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Passage-Dependent		32.43	50.00	21.74	42.31	86.96	45.00	46.43	73.68	44.74	83.33
Passage-Independent		67.57	50.00	78.26	57.69	13.04	55.00	53.57	26.32	55.26	16.67
Total % of Items		100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Thus a student who wants a relatively quick examination with a fairly large proportion of multiple-choice items should focus on taking private university examinations, or, even more so, on the Center examination.

5. What skills are measured on these entrance examinations?

As discussed above in the *Definitions* section, there are differences in the skills necessary to answer discrete-point test items and integrative ones, as well as between receptive test items and productive ones. We also argued that translation is probably too difficult, demanding, and specialized a skill to require of students who have only studied a language through junior and senior high school. Tables 5A and 5B again present results for these different item categories for the private and public universities, respectively.

Among the private universities (Table 5A), the discrete-point items dominated all of the tests except for KEIO (which is 50 percent translation, 37.5 percent discrete-point, and 12.5 percent integrative). However, fewer discrete-point items appeared on the public university exams (Table 5B), which appear to vary more in terms of the proportions of these three item categories.

For private universities, about the same pattern occurred as in the discrete-point results, that is, receptive items dominated, probably because discrete-point items tend to be receptive (though some can be discrete-point and productive, e.g., single discrete-point fill-in items which require students to produce language). Table 5B indicates that on the public university tests the proportions of receptive items are generally smaller and that the proportions of receptive, productive, and translation items vary more from university to university than among the private universities.

Table 6 compares the averages for private and public universities with the Center examination and the total for all 21 universities. This table reinforces the observations made above that the private university examinations were more likely to use discrete-point receptive items than the public ones. However, though the public university examinations used more integrative items (29.11 percent on the average), they also relied heavily on translation items (26.29 percent).

Table 6 also indicates that similar patterns emerge for the receptive, productive, and translation categories of items. It appears that examinees are not being required to produce much language at all, even written language. They certainly are not being required to produce any spoken English on any of these tests.

The skill of listening, which is being heavily promoted in the new Ministry of Education guidelines (Monbusho, 1989), appears to be im-

Table 6: Categories of Item Types Summarized by University Type*

Item Category	Private	Public	Center	Total
<i>Frequencies</i>				
Discrete-Point	39.80	12.30	48	27.10
Integrative	1.80	6.80	5	4.33
Translation	2.10	5.90	0	3.81
Total No. of Items	43.70	25.00	53	35.24
Receptive	41.20	9.80	53	26.81
Productive	0.40	9.30	0	4.62
Translation	2.10	5.90	0	3.81
Total No. of Items	43.70	25.00	53	35.24
Passage-Dependent	15.30	12.30	15	13.86
Passage-Independent	28.40	12.70	38	21.38
Total No. of Items	43.70	25.00	53	35.24
<i>Percentages</i>				
Discrete-Point	86.08	44.61	90.57	66.54
Integrative	4.68	29.10	9.43	16.54
Translation	9.24	26.29	0.00	16.92
Total % of Items	100.00	100.00	100.00	100.00
Receptive	88.63	37.19	100.00	64.68
Productive	2.13	36.52	0.00	18.40
Translation	9.24	26.29	0.00	16.92
Total % of Items	100.00	100.00	100.00	100.00
Passage-Dependent	41.71	52.66	28.30	46.29
Passage-Independent	58.29	47.34	71.70	53.71
Total % of Items	100.00	100.00	100.00	100.00

* All statistics for private and public universities as well as total are averages.

portant at only some of the universities. Tables 3A and 3B indicate that six of the 21 examinations included listening comprehension items of some sort: AOYAMA, TSUDA, HITOTSU, KYOTO, TOKYO, and TYOUPS. [note that TOKYO had listening comprehension items, but they were not available in Koukou-Eigo Kenkyuu (1993b) and therefore could not be included.]

Discussion

What do the results presented in the previous section mean to teachers and students of EFL?

Readability. The Flesch-Kincaid and Fog readability indices indicated that the passages would be appropriate for native speakers ranging from eighth grade (about 13 years old) to thirteenth or even fifteenth grade level (university age). The passages at the upper end of these scales (most notably those at KEIO, KYOTO, TOKYO, and YOKOHAMA) are probably particularly difficult reading material for EFL students at the level of those taking the examinations. None of the passages used in these tests could be described as simplified texts based on the readability levels alone. In general terms, the public universities had more difficult passages (one full grade level higher on the Flesch-Kincaid and Fog scales) than the private universities and the Center examination passage was easier than those on the private or public examinations.

For those of us in the EFL world, this means that overall, by third year, we should be pushing high school students into reading passages that are relatively difficult—and certainly not simplified passages. Indeed, those aiming to enter KEIO, KYOTO, TOKYO, or YOKOHAMA should be prepared to read university level texts in English with good comprehension.

Item types. The analysis of item types (see Tables 3A, 3B, and 4) indicated that EFL/ESL students should be equipped to grapple with a variety of types of test items, as well as with considerable variation within each type. In short, examinees should be prepared for considerable variation in the types of items and tasks that they will be facing, as well as for frequent changes in instructions.

As some universities placed great emphasis on multiple-choice items, while other universities emphasized translation, and still others, listening (both multiple-choice and fill-in), those preparing for different universities can expect to have to prepare for different types of items, indeed, for different types of tests.

The considerable differences in test length, both in time and in number of items, can not only be important to individual test takers, but can also have important policy level implications because there is a direct relationship between test length and test reliability: longer tests are generally more reliable than shorter tests if all other factors are held constant (see Brown, in press).

Conclusions

Perhaps the single most important fact about these very competitive entrance examinations is that the results are used to make decisions about students' lives—important decisions. As such, the examinations must be of the highest quality if they are to be fair to the students.

Generally speaking, the individual items on these tests are reasonably well written from the point of view of proofreading and quality of language. We found very few malapropisms, typographical errors, unintentional grammatical errors, etc. However, that does not mean that the tests are perfect or that they are perfectly fair to all who take them, year after year.

Many test questions are based on all or portions of passages, many of which are difficult. Consequently, the ability to answer the questions will depend to some degree on one's ability to deal with relatively high level language, almost surely above the level of the simplified texts that are often used for pedagogical purposes. This, in and of itself, is not a major problem. After all, examination scores ought to have some relationship to English language reading ability.

However, when we recognize that the ability to answer many of the questions may depend on knowledge of the topics involved in the passages, there is a much more serious threat to the reliability and validity of these examinations. In other words, chance knowledge of a particular topic and the vocabulary that goes with it might determine to some degree whether or not students are accepted into this or that university.

Western teachers are sometimes cynical about the Japanese obsession with sending their children to a *juku* (cram school) or *yobiko* (test-coaching school) to prepare for major entrance examinations. Indeed, before analyzing these results, we too were skeptical of the value of *juku* preparation. However, from our analysis of the tests, we have concluded that such preparation may be useful, and even advisable. There is considerable variation in the types of items used on these tests, especially in the public university examinations. Such variation means that examinees are often reading new directions and shifting gears in the kinds of tasks that they are doing, and, as the ability to understand directions and shift gears in a test is related to testwiseness, it is crucial that students acquire such abilities. One way to acquire this testwiseness is by attending a *juku* or *yobiko*. Given the competitive nature of these examinations, with only a very small percentage "passing" and being accepted to these universities, it may be that testwiseness, or the ability to take tests in general, is as important or perhaps more important than actual abilities in English.

In addition, we must recognize the effects of different item types on what is being tested. The private universities predominantly use discrete-point receptive items. This means that in effect they are endorsing a discrete-point and receptive view of language teaching. However, before being too critical, one must admit that the predominant North American

examination, the TOEFL, is also discrete-point in nature. However, it must also be recognized that the TOEFL 2000 project is aimed at changing the TOEFL into a more communicative, integrative, task-oriented examination. Perhaps Japanese universities will begin to move their examinations in similar directions.

The public university reliance on translation items might be said to be out of date because translation was abandoned years ago in ESL instruction. (The factors behind their retention entailing a conviction that "meaning" is expressible only in the L1 [Japanese] are beyond the scope of this discussion.)

There is also a contradiction between what is tested in these examinations and the Ministry of Education (Monbusho, 1989) aural/oral guidelines implemented in April 1994. It will be interesting to see how the structure and nature of the entrance examination items change over the next several years to reflect the new emphasis on aural/oral skills.

The universities in this study have openly provided the examinations for publication. Such publication of tests is useful because it allows for public scrutiny. However, that is not enough. These institutions are also responsible for making sure that their tests are practical, reliable, and valid. There are good reasons to believe that the examinations investigated in this study may be weak in all three areas. For instance, the tests may not be 100 percent practical because some item types are very difficult to score, because item types change frequently, and because many of the items are passage and topic dependent. In addition, many of the tests are relatively short, and therefore often yield unreliable results. Finally, the types of items used on these tests may be so out of synch with current language teaching theory and practice that there are serious reasons to question their content and construct validity.

How can these problems be avoided? We strongly suggest that these universities and the language professionals who write the tests consider following the *Standards for Educational and Psychological Testing* (CDSEPT, 1985). In the United States, tests became practical, reliable, and valid only when students and their parents filed lawsuits against the various organizations that developed tests. The *Standards for Educational and Psychological Testing* then appeared and began to be applied. In addition, *Buros Mental Measurements Yearbook* (e.g., Kramer & Conoley, 1992), which is a collection of reviews of published tests, serves as a critical watchdog on all published tests in the United States. Both the *Standards* and *Buros* are designed to keep test developers honest. Similar institutions might prove similarly useful in Japan. The point is that the entrance examinations in Japan are far too important to

be left entirely up to the test designers. Professors must be made accountable, perhaps for the first time, for the important admissions decisions that they make because those decisions so profoundly affect young people's lives.

Note

1. We would like to thank Ryutaro Yamashita for his help in entering the reading passages into computer files.

James Dean Brown (J.D.) is on the faculty of the Department of ESL at the University of Hawaii and often teaches at Temple University Japan. He has published numerous articles on language testing and curriculum development, and three books on: reading statistical studies (Cambridge), language curriculum development (Heinle & Heinle), and language testing (Prentice-Hall).

Sayoko Okada Yamashita teaches in the Japanese Language Program at the International Christian University in Tokyo. She has published a number of articles and books on language teaching and is currently a doctoral candidate in the TESOL Program at Temple University Japan.

References

- Amano, I. (1990). *Education and examination in modern Japan*. Tokyo: Tokyo University Press. (Translated into English by K. & F. Cummings.)
- Borland. (1991). *QuattroPro* (version 3.0). Scotts Valley, CA: Borland International.
- Brown, J.D. (in press). *Testing in language programs*. New York: Prentice-Hall.
- Carroll, J.B. (1972). Fundamental considerations in testing for English language proficiency of foreign students. In H.B. Allen & R.N. Campbell, *Teaching English as a second language: A book of readings* (2nd ed.). New York: McGraw-Hill.
- CDSEPT (Committee to Develop Standards for Educational and Psychological Testing). (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Cummings, W.K. (1980). *Education and equality in Japan*. Princeton, NJ: Princeton University Press.
- Duke, B. (1986). *The Japanese school: Lessons for industrial America*. New York: Praeger.
- Flesch, R. (1974). *The art of readable writing*. New York: Harper & Row.
- Frost, P. (1991). Examination Hell. In E.R. Beauchamp, *Windows on Japanese education* (pp. 291-305). New York: Greenwood Press.

- Hinofotis, F.B. (1981). Perspectives on language testing: Past, present and future. *Nagoya Gakutn Daigaku Gaikokugo Kyoiku Ktyo*, 4, 51-59.
- Koukou-Eigo Kenkyuu. (1993a). '93 *Kokukouritsu daigaku-ben: Eigomondai no tetteteiki kenkyuu* ('93 national and public universities: English test problem research). Tokyo: Kenkyusha.
- Koukou-Eigo Kenkyuu. (1993b). '93 *Sbtritsu daigaku-ben: Eigomondai no tetteteiki kenkyuu* ('93 private universities: English test problem research). Tokyo: Kenkyusha.
- Kramer, J.J. and Conoley, J.C. (Eds.). *The eleventh mental measurements year-book*. The Buros Institute of Mental Measurements, Lincoln, NE: University of Nebraska Press.
- Monbusho. (1989). *Chuu gakkoo gakushuu sbidoo yooryoo* (Course of study for junior high school). Tokyo: Monbusho.
- Que Software. (1990). *RightWriter: Intelligent grammar checker* (version 4.0). Sarasota, FL: Que Software.
- Shimahara, N. (1991) Examination rituals and group life. In B. Finkelstein, A.E. Imamura, J.J. Tobin (Eds.), *Transcending stereotypes: Discovering Japanese culture and education* (pp. 126-134). Yarmouth, ME: Intercultural Press.
- Spolsky, B. (1978). Introduction: Linguists and language testers. In B. Spolsky (Ed.), *Advances in language testing series: 2*. Arlington, VA: Center for Applied Linguistics.
- Tsukada, M. (1991). In B. Finkelstein, A.E. Imamura, J.J. Tobin (Eds.), *Transcending stereotypes: Discovering Japanese culture and education* (pp. 178-182). Yarmouth, ME: Intercultural Press.
- WordPerfect. (1988). *WordPerfect* (version 5.1). Orem, UT: WordPerfect Corp.

(Received August 4, 1994)