Bulletin of Hokuriku University
Vol. 18 (1995)
pp. 61 ~ 75

A Comparative Analysis of Entrance Examinations

John D. Dennis *

Received October 31, 1995

This paper will survey the results of a statistical analysis of the entrance examinations used by Hokuriku University. I hope that readers of this paper will gain a better understanding of these examinations in two ways. First, this understanding may be based simply on the merits and deficiencies of the entrance examinations used at Hokuriku University, without regard to the data that are available from other universities in Japan. As a follow-up step, the reader may wish to contrast the data available for Hokuriku University with similar statistical data that has been published for a number of other universities in Japan.

Initially, statistical data drawn from past entrance examinations given by Hokuriku University will be analyzed in terms of common readability scores to show the consistency, uniformity, and degree of difficulty of the reading passages on these examinations. Next, this data will be compared and contrasted to equivalent data that has been published for the examinations administered by 20 other well-known domestic universities.

Brown and Yamashita (1995) studied the entrance examinations that 20 Japanese universities administered in 1993. They divided their analysis between 10 public and 10 private universities. The schools used in their study are listed here. The public universities are: (1) Hitotsubashi University (Hitotsu), (2) Hokkaido University, (3) Kyoto University, (4) Kyushu University, (5) Nagoya University, (6) Osaka University, (7) University of Tokyo, (8) Tokyo University of Foreign Studies, (9) Tokyo Metropolitan University, and (10) Yokohama City University. The private universities are: (1) Aoyama Gakuin University, (2) Doshisha University, (3) Keio University, (4) Kansai Gaidai (Foreign Languages) University, (5) Kansai University, (6) Kyoto University of Foreign Studies, (7) Rikkyo University, (8) Sophia University, (9) Tsuda University, and (10) Waseda University.

The schools in the Brown and Yamashita study were chosen for two reasons: due to their prestige, and in order to reflect a reasonable geographic distribution across Japan. The intent of their research was to display the characteristics of these examinations so that they could draw conclusions about the general system of entrance examinations in Japan. In addition to the universities listed above they also included an analysis of the "Daigaku Nyuushi Center" examination. Traditionally, universities in Japan have almost always prepared their own

* 英米語学科
Department of English

entrance examinations, but this "Center" exam has gained a certain respect due to the number of universities that make use of it. It is similar in function to the testing system used in the United States in that it is developed and administered by an institution that is separate and distinct from the universities that use the results of the test. This test is important in Japan because it is administered nationwide, and it is used by a number of universities, either exclusively or in conjunction with their own exams.

Brown and Yamashita attempted to study the general level of difficulty of these examinations; differences in the level of difficulty between examinations; the types of items (questions) used, how varied these item-types were, and difference in test length; and finally, what types of skills were measured on these examinations. Generally, their study aims to provide English teachers with a snapshot of the examinations given in Japan so that they may more successfully prepare their students for the exams. Here, however, I would like to focus instead on how this data might be used as tool by those preparing tests in order to produce tests that are, on the one hand, more practical, uniform and fair, and on the other hand, more reliable and valid.

I have attempted to replicate a portion of the procedures used by Brown and Yamashita in their study, and the data displayed here for all universities other than Hokuriku University is theirs. The types of questions, the purpose of different question types, the number of answer options given to students, the language(s) used in a question, and the tasks involved were all cataloged and analyzed by Brown and Yamashita. Although I have presented only a portion of their statistics, this paper does takes a major step toward a thorough analysis of Hokuriku University's entrance examinations and a comparison of those results to the data produced by Brown and Yamashita.

The statistical data for the reading passages used in Hokuriku University's entrance examinations was produced using a computer program called *Grammatik Mac*™ (Reference International Software, 1990). This data consists of such count categories as the number of passages, the number of words, the number of unique words (counting all words but counting repeated words only once, no matter how often they may recur), the percentage of unique words (type-token ratio), the number of sentences, the number of syllables per word, and the number of words per sentence.

The number of passages is important because it shows the variety of topics on a test. Most university entrance exams have a variety of readings, while others have just one (Keio). Longer readings would seem to require deeper, more complete knowledge of a specific topic, while tests that use a variety of readings may require a broader range of knowledge, without any of it being too complete. The number of words, either in total or the average number of words per passage, also indicates something about the kind of reading skills needed. More words means more reading, and perhaps a broader and deeper knowledge of topics along with it. More words may also mean a school values reading skills more highly than other skills such as translation, listening, vocabulary, or grammar. The number of unique words is the

number of different words used. This shows the variety of vocabulary used in a reading. The type-token ratio is the percentage of unique words, which is calculated by dividing the number of unique words by the total number of words. The number of syllables per word reflects the difficulty of the vocabulary used in a reading. A higher figure indicates a reading is more difficult. A higher number of words per sentence indicates that a reading is syntactically or grammatically more complex, and that it will require better language skills to understand.

Besides these categorical analyses three readability indexes were calculated. The first of these is the Flesch Reading Ease score, the second is the Flesch-Kincaid Grade Level score, and the third is the Gunning's Fog Index. While readability scores can vary (they are based on different formulas), and while statistical indicators like these are open to challenge, these indexes have been shown to reflect the level of difficulty of various texts. I hope that by including three such measures any challenges to a statistical analysis of reading difficulty will be minimal and readers will have a chance to base their judgements on a broad range of indicators rather than relying on just one index. Instead of simply providing certain numbers and expecting readers to take them on faith, it might be useful to examine how these different indexes arrive at their scores.

First, the formula for the Flesch-Kincaid Grade Level score is: (0.39) x (average number of words per sentence) + (11.8) x (average number of syllables per word) = total - 15.59 = GRADE LEVEL. Since the Flesch-Kincaid score is expressed in terms of grade level, a higher grade level naturally means that a text is more difficult. Levels of 6th to 10th grade are most common, and are considered to be most effective for communication among native speakers of English. While scores that are higher may seem to be more scholarly or learned, such texts may instead simply be more difficult to understand.

Second, the formula for Gunning's Fog Index is: (average number of words per sentence) + (number of words of 3 syllables or more) = total - 0.4 = Fog Index. As with the Flesch-Kincaid score, the Fog Index score is intended to measure the approximate grade level a reader must have achieved to easily read and understand the text concerned. Readers may want to note that Fog Index scores are typically higher than Flesch-Kincaid scores. Again, a higher score means a text is more difficult to read.

Third, the formula for the Flesch Reading Ease score is: 206.835 - {1.015 x (average sentence length) + 0.846 x (number of syllables per hundred words)} = Flesch Reading Ease score. This score is on a scale of 0-100. Here, in contrast to the above two indexes, a lower score means a text is more difficult to read, a higher score means it is easier. A score of 90-100 is considered to be very easy, about 4th grade level; a score of 60-70 is considered to be standard, about 7th-8th grade level; and a score of 30 or below is considered to be very difficult, from college level on up.

The reading passages from Hokuriku University's entrance examinations for two successive years (1994 and 1995) were entered into a computer and analyzed in terms of these statistics using the *Grammatik Mac*™ program. Due to the nature of the entrance examination

system at Hokuriku University an opportunity exists to study the difficulty of the reading passages in a variety of ways. Six tests are offered each year to potential candidates for matriculation, a system which is quite different from the testing systems of any of the public or private universities listed earlier. Additionally, since each test has multiple reading passages, the tests in this program can be examined from several perspectives. Thus, (1) passages can be analyzed individually, (2) the passages on any one of the 12 tests can be lumped together and viewed as a single unit and analyzed to show test-specific results, (3) tests from a given year can be lumped together and the results of the analysis averaged to render a picture of a typical test for that year, or (4) all passages can be lumped together and the results of the analysis averaged to produce data for a model "test" against which comparisons can be made.

Here, I have chosen to present the data for the first three of these perspectives. This data is contained in the appendixes at the end of this paper. First, Appendix 1 presents an analysis of the individual reading passages for the six examinations used in 1994. Each test and passage has been coded in the following way: test year (in this case all passages are from the 1994 series), letter (indicating separate test forms), and the number of the reading passage (one number for each passage on a test form). Together with the averaged results contained in Appendixes 2 and 3, this data could help an entrance examination committee identify specific passages that are either too easy or too difficult, so that such passages could be edited, eliminated, or complemented with other passages to balance the overall level of difficulty on a test or series of tests. Due to the type of testing program at Hokuriku University, this level of analysis quickly produces large amounts of data, but this will be necessary if we hope to study, and then later control or manipulate the difficulty levels of our tests.

These multiple perspectives allow an analysis that will show whether or not there is any variation in the level of difficulty of the reading passages on Hokuriku University's entrance examinations. Initially, we must simply ask if there is variation and how it can best be described. If there is variation, we need to ask whether it is excessive or not. In other words, is there too much variation in the level of difficulty, or does it fall within normal limits? Finally, we might ask whether variation in the level of difficulty of reading passages is desirable, when considering program goals and when designing appropriate tests. Should there be variation? If so, how much? Will variation in the level of difficulty of reading passages give better (or worse) numerical results, perhaps allowing us to more easily distinguish candidates who should be accepted for study?

It is also important to ask where variation in the level of difficulty may or may not be desirable. At the lowest level, within a specific test, difficulty may vary from passage to passage. Presumably, more variation in the level of difficulty at this point would yield a broader range of scores than if there were less variation. A test with minimal variation in the level of difficulty from passage to passage would probably define a threshold: candidates whose competence was above the threshold would do quite well on the test while candidates

whose competence was below the threshold would do poorly. This would be fine if the threshold was at the proper level of language competency. Then the proper candidates would be adequately identified. However, if the threshold is at the wrong level of competency there may be problems. If the threshold is too high candidates with a lower level of skill would be poorly differentiated, making the selection of the best candidates among these relatively unfair–more a toss of the dice than based on reliable test results. If the threshold is too low candidates with a lower level of competency may be grouped in an unreliable way with those whose skills are much better. Therefore, unless a test has been properly designed it would seem to be safer and more equitable to insure a certain amount of variation in the level of difficulty on a test.

While a reasonable variation in the level of difficulty within a test may be preferable, variation between tests should obviously be avoided. The most prestigious universities offer a single entrance examination each year and thus do not have to worry about inter-test variation in the level of difficulty. However, as stated above Hokuriku University offers six different examination dates (and six different tests) each year. While this is convenient for applicants, care must be taken to insure as much uniformity as possible when preparing these tests.

Thus, the statistics described above may be quite useful for making comparisons between multiple test forms that are supposed to be equivalent. These statistics permit an analysis of past test forms, allowing us to judge whether they are acceptably uniform or not. They also permit an analysis of tests in preparation, allowing us to minimize differences between test forms and consequently administer uniform tests that are fair and equitable to the candidates taking them. Such tests will also more clearly identify the most suitable candidates for matriculation and will bolster the integrity of the program they are entering. Given that we have to accept students of mixed ability, we should certainly try to select those with the best possible skills, which will make for a better program and better progress of the students in their studies.

This inter-test level of analysis is the most important for identifying possible problems in test design that will affect both students and the program they enter. Unfortunately, to my knowledge, Hokuriku University's tests have never been analyzed in this way. Uniformity in the level of difficulty between tests has never been studied (nor has the numerical data that is produced by student scores). This paper is a first step towards the kind of attention that should be paid to this issue. The importance of this kind of care in the design and preparation of Hokuriku University's entrance examinations is natural and should be ongoing if we are to meet normally accepted standards of test administration. Nothing more than simple statistics are needed to begin this type of monitoring.

We should be especially cautious because tests of uneven difficulty that are offered on successive days may unfairly discriminate among students. Presenting a system of examinations that is convenient to students may have its advantages, but if the test forms are not equivalent in terms of difficulty, then admission becomes a game of chance rather than

skill. Students who opt to sit for more than one examination may find their chances of acceptance increased, but may also be left with the perception that luck determined their fate more than skill. Stability, predictability, and uniformity will inculcate respect. Care should be taken that the impression of variation in the level of difficulty does not take root among students who are preparing for examinations, or among their teachers. The impression of chaos in an entrance examination system, something that often determines a person's fate, should be avoided at all costs.

The final level of variation in examinations is not so crucial. This is the macro level: variation in the level of difficulty from year to year. From this vantage point there may be variation in the level of difficulty, and it may be intended or unintended. Even unintended variation at this level will have a much less dramatic effect on students and the program they are entering, since there is comparatively little chance of fairness becoming an issue, and negative impressions will be minimal. On the other hand, year to year variation in difficulty and other aspects of test design may be the result of intentional planning, such as a curriculum change or a change in the goals that underlie it. It is important to note that in their study Brown and Yamashita collected data for a large number of schools but for only one year–1993. Therefore, we don't know the degree of variation in the level of difficulty of these schools from year to year. As I will discuss below, Hokuriku University's entrance examinations do show some year to year variation, and even from test to test within the same year. However, this may simply be visible because the results have been displayed here. Also, since we give six tests per year rather than the single examination that many of the most prestigious universities do, test-to-test consistency and reliability is of more concern. If the examinations for the schools in Brown and Yamashita's work were studied over the course of multiple years, there could be either more or less variation in the level of difficulty than is exhibited by Hokuriku University. At this point it is difficult of impossible to say what "normal" variation might be. Resolving such a question, while important, is beyond the scope of this paper.

Next, I would like to discuss the data shown in Appendix 2a and Appendix 2b. These tables show the readability statistics for twelve separate tests used by Hokuriku University as entrance examinations. Appendix 2a contains the data for six tests administered in 1994 and Appendix 2b contains the data for six tests administered in 1995. These tests have been coded so that they can be identified by year and letter. The year, either 94 or 95, shows the year that students taking a test would matriculate, and the letters A through F identify the six individual test forms used during that year. For purposes of comparison I have also included the yearly averages for the six tests offered during each year.

By comparing these twelve tests it can be seen that, based on readability statistics for the reading passages that they contain, some tests certainly appear to be more difficult than others. I have identified the easiest and hardest tests overall and the easiest and hardest tests for each year. I have based these judgements almost exclusively on the Flesch, Flesch-

Kincaid, and Fog scores. While the other statistics tend to support these judgements there is some variation. For example, I have identified test 94D as both the hardest test during the 1994 test year and also as the hardest test overall. The Flesch, Flesch-Kincaid, and Fog scores clearly indicate this, but note that test 94E has a higher type-token ratio, i.e., a higher percentage of unique words. Test 94D clearly has the longest sentences by a wide margin (an average of 28.4 words per sentence) which contributes to its high degree of difficulty ratings in other categories.

Selecting the easiest test from among these twelve is not such a distinct choice. I have selected test 94F as the both the easiest test during 1994 and also as the easiest test overall. However, there are three other tests that appear close to this one in level of difficulty. Test 95A, the easiest test of the 1995 test year, comes in a close second for easiest overall. Also, note that the statistics for tests 95B and 95E indicate that their reading passages are also relatively easy.

Next, note that the average scores for the 1994 test year indicate a uniformly more difficult set of readings than in 1995. First, the tests administered during 1994 have more readings (four rather than three), they have a higher percentage of unique words (type-token ratio), they average more syllables per word (1.58 versus 1.45), they tend to have longer sentences, and the three readability scores (Flesch, Flesch-Kincaid, and Fog) are all higher. Clearly, with this much agreement among various statistical measures, it is easy to conclude that the reading passages used during the 1994 test year are more difficult than those used during the 1995 test year.

It is also important to notice the variation in the level of difficulty of tests within each test year. The range of variation in the level of difficulty of the reading passages in different tests is quite different from 1994 to 1995. Individual tests in the 1995 test year all stay relatively close to that year's average scores. The range of variation from high score to low in any of the statistics used here is relatively small. However, individual tests in the 1994 test year show more variation from that year's average scores, and there is a large difference in difficulty level between the most difficult and easiest tests. This can be easily seen by contrasting tests 94D and 94F, the tests that I identified above as the easiest and most difficult tests over this two year span of time. As a general goal for test design, tests which purport to test the same skills and abilities should be as uniform as possible. Thus we can conclude that the lesser amount of variation in readability scores in 1995 indicates a better set of tests than the comparatively large variation in scores that we can see for the tests used in 1994.

Now I would like to turn to the data shown in Appendix 3a and 3b. Appendix 3a allows us to see the similarities and differences in reading passage statistics for Hokuriku University and some other selected schools. As described above, Brown and Yamashita examined reading passages from the exams of ten public and ten private universities. I have chosen four of these schools to compare to Hokuriku University's scores. I have placed the scores for tests from Hokuriku University that have the easiest and hardest reading passages alongside

the public and private schools that, respectively, have the easiest and most difficult reading passages. Clearly, the reading passages in Hokuriku University's easiest test seem to be on par with those used by Kansai University and Hitotsubashi University. In fact the Flesch, Flesch-Kincaid, and Fog scores are all exceptionally close.

However, when we look at the tests with the most difficult reading passages there is less agreement. Note that Hokuriku University's most difficult reading passage scores indicate greater difficulty than the scores available for either of the schools included in the Brown and Yamashita study. It is striking that on the one hand, Hokuriku University's readability scores for its easiest test are on par with scores from schools that have the easiest reading passages, while on the other hand the scores for Hokuriku University's most difficult test show that these readings are even more difficult than readings used in the tests for Keio University or Yokohama City University. When taken together with the point that both the easiest and most difficult reading passages used by Hokuriku University were offered during the same test year this would appear to show that there may be too much variation in the level of difficulty in the reading passages used here.

The difference becomes acceptable, however, when Hokuriku University's scores are averaged for 1994 and 1995 and then compared to the averages for public and private universities, and also with the "Center" examination. Appendix 3b shows that the "Center" examination has the easiest readability scores, and that the average scores for Hokuriku University are reasonably close to the average scores published by Brown and Yamashita in their study for public and private universities. In fact, if Hokuriku University's average scores are taken to define a range of difficulty, then it can be seen that the average scores for public and private universities tend to fall within this range. For example, Hokuriku University's average Flesch scores are 57 and 66, and the average scores for public and private universities fall within this range. The same can be said for the Flesch-Kincaid, Fog, and syllables per word statistics. Note that for these same statistics the "Center" examination falls outside of and on the easier side of this range of difficulty.

A cautionary word is important at this point. Hokuriku University is unusual in offering six different examination dates each year. Thus there is a great opportunity to observe and examine variation in the level of difficulty of the reading passages that it uses. On the other hand, Brown and Yamashita's study examined single examinations that were offered by schools that typically offer only one such exam each year. Also, they studied exams from their selected schools for only one year, rather than pursuing a longitudinal study where they would be able to see change or variation over the course of time. Thus it is impossible to say whether the schools in the Brown and Yamashita study exhibit more or less variation in the level of difficulty of their reading passages than does Hokuriku University. At this point in time it is not possible to say what an acceptable or normal range of variation in the level of difficulty of reading passages might be.

Another possible direction that this kind of comparison of the levels of difficulty in

reading passages might take is to look at the proficiency tests offered by different companies to the general public. It would be particularly interesting to examine the reading passages of tests such as the TOEFL, TOEIC and STEP tests. Regarding these tests, we need to ask what are their levels of difficulty are in terms of these readability statistics and how the results of a similar analysis of their reading passages would compare to the data available here. While answering this type of question is truly beyond the scope of this paper, it is possible to elaborate on the directions that this kind of analysis might take. I have not attempted to analyze passages lifted from these sources, but examining retired reading material from these tests in this way seems very appealing.

Indeed, just thinking about calculating readability statistics for passages on those tests produces a good model for conceiving of the depth and quality of information readability statistics are producing. Though they may accurately identify difficult readings, and although they may be strongly correlated with human judgements of difficulty, the kind of statistics used here do not explain "why" people perceive certain passages, questions, or tests as difficult. Some of the readability statistics described here stand in a distant and abstract relationship to some human judgements about the difficulty of certain tests.

For example, if asked to complete the thought: "That test's hard because...," examinees might answer with some of the following reasons, which may be classified according to type. One group of reasons they put forth may have little to do with readability statistics. Examples of responses that would fall into this category are: "It's too long," "I didn't know anything about the topics, they were unusual," or "The readings were okay but the questions were too hard." Another group of reasons that examinees may use to explain test difficulty might be related to other testing concerns. Statements such as: "I didn't have enough time," "I don't like multiple choice questions," or "I didn't understand the directions," would fall into this category.

A final group of reasons might be related to some of the factors used by these statistics: "There were a lot of words I didn't know," "There were too many difficult words," or "The sentences were really long." Reasons like this are informal expressions of statistics such as the number of unique words per passage, the type-token ratio, number of syllables per word, or the number of words per sentence. In spite of this variety of responses I doubt anyone would offer an explanation at all close to the complicated formulas that I described earlier (the Flesch, Flesch-Kincaid, and Fog scores). These statistics, though accurate, do not resemble the more intuitive accounts that real people would provide. At least in part, this is because they are combinations of these accounts, which are used together in specific ways to provide a more reliable picture of the levels of passages.

Brown and Yamashita's study focused on how to prepare students for these exams, but here I want to develop a different perspective: how to prepare exams for prospective students. In order to do this I need to explore the relationship between entrance testing and curriculum. Generally stated, the concept of program and goals that underlies curriculum should be the

guiding light for test design. Good program conceptualization should have two results: the first of these is good curriculum design and its subsequent operation, and the second is to clearly identify a good set of criteria, standards, and goals for designing, planning, and writing entrance examinations.

If this relationship between concept, curriculum, and exam design is cultivated and carried through, then entrance exams will be more stable. Good criteria for exam design will produce more continuity and predictability in terms of test content and difficulty. Thus stability or predictability may indicate a better or clearer underlying program concept and more detailed curriculum goals, besides the administration of these–good planning and the follow-through to make it happen.

However, regardless of the goals of a program, and however much these goals help define its curriculum and entrance exam, any school would probably want to standardize the difficulty of the tests that they use. This standardization could begin with a restricted scope, perhaps by identifying problem areas or by prioritizing the sections of a test that need attention. For example, faculty might identify the reading section of a test for improvement, subsections of which are the reading passages themselves and then questions based on those passages. The readability statistics used in this paper may be useful for gauging the difficulty or uniformity of possible reading passages and enable faculty to choose the best variety of readings. Perhaps they would set thresholds to the effect that passages should fall within minimum and maximum levels of difficulty.

It might be said that the pressure for curriculum change is an explicit, obvious need; but that the pressure for studying and analyzing the entrance exam is an implicit one. Reports are available for the total number of schools in Japan and also for the number of those that have recently changed their curriculums. In contrast to a more wide-spread awareness of curriculum change, similar news about the entrance examinations used by schools is comparatively meager. This may be because most schools only change curriculum, without really changing their program's underlying concept or their entrance examination practices. How many schools that have changed curriculum have also had the degree of insight and the perseverance to adapt their entrance exams to their changed curriculum requirements? Curriculum change and entrance exam change are the two practical aspects of a change in the underlying conceptualization of a program. The follow-through from curriculum change to testing should be given a high priority.

This lack of attention to entrance examinations is unfortunate since modernizing and renewing entrance exam standards should also have some public relations value. This type of value can be significant, and it might be possible to make more of these exams than first glance would reveal. This value seems to have been partially recognized in that some schools have been creative in opening multiple and alternative paths to matriculation. Recognition and coordination of the testing-curriculum relationship can create the best academic world for matriculating students. It also selects the students that are best-suited to the program and

goals that have been laid out.

Finally, this kind of readability analysis could be applied in two different ways: passively or actively. In this paper we have seen the passive role that readability statistics can play. They can be used to analyze and compare different tests, changes in testing policy on a year-to-year basis, and to display the difficulty level of individual passages on a test. However, using these statistics in an active way might also be attempted. It may be possible to use this kind of analysis to "groom" a test–to adjust or adapt its level of difficult to conform to certain guidelines. Simplifying texts, by substituting shorter words, altering or shortening sentences, seems easier to grasp than its opposite: somehow complicating a text, making it more complex and difficult. Simplification seems natural and likely to be an easier process than "complicating" or "complexifying" a specific passage. Rather than using it as a basis for editing, this kind of analysis would more easily be used to help in the selection of readings during test construction.

Conclusions

One of the biggest questions surrounding this kind of statistical analysis is whether it is really accurate. Does it really reflect difficulty of a test? Or does it influence difficulty only partially, in that the kind and quality of the questions and tasks may also be important? Or is difficulty of a test only a minor function of the passages themselves and more a function of the questions and tasks based on them? Also, this kind of analysis focuses directly and exclusively on reading passages, while ignoring test items not based on reading passages, such as translation, reorderings, shorter sentence-based tasks and so on. Also, the effects of time pressure are almost impossible to gauge. One possibility alluded to earlier was to analyze reading passages used in the TOEFL or TOEIC examinations. By gaining an understanding of the range of variation in level of difficulty on these tests we may begin to define what level or what range of difficulty is to be targeted. Should we opt not to use readings that cross the threshold of being "too difficult" or "too easy?" Should we try to use readings all of which have the same general level of difficulty? Or should we try to develop tests with multiple reading passages that vary in their level of difficulty?

Though important, answering these questions would perhaps be more difficult than reading any of the passages analyzed in this paper or in other similar studies. Part of this problem is understanding in what terms people perceive difficulty. The types of tasks and the factors involved in judging overall difficulty of a test are extensive, and exploring even a small portion of these would require considerable energy and effort. Part of the answer to these questions lies in the response rates of students who are taking these exams. If we assume that any group of students who are taking a given test are equal to any other taking another, supposedly equivalent test, then we may be able to correlate their scores on different test forms with the kind of statistics discussed in this paper. These student scores would help us more clearly understand student perception of and response to levels of difficulty of reading

passages as they are perceived in the tests themselves, rather than as singular, isolated reading passages. Also, it would be interesting to look at students who have matriculated and who are now successful. Then we can ask how statistical analyses such as these can help us attract and admit more of these same kind of students.

References

Brown, J. D. & Yamashita, S. O. (1995). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal, 17,* 7-30.

Brown, J. D. (1995a). Differences between norm-referenced and criterion-referenced tests. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 12-19). Tokyo, Japan: Japan Association for Language Teaching.

Brown, J. D. (1995b). Developing norm-referenced tests for program-level decision-making. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 40-47). Tokyo, Japan: Japan Association for Language Teaching.

Childs, M. (1995). Good and bad uses of TOEIC by Japanese companies. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 66-75). Tokyo, Japan: Japan Association for Language Teaching.

Clankie, Shawn. (1995). An introduction to commercial English tests in Japan. *The Language Teacher, 19* (6) 8-10.

Gilfert, Susan. (1995). A comparison of TOEFL and TOEIC. In J. D. Brown & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 76-85). Tokyo, Japan: Japan Association for Language Teaching.

Reference Software International. (1990). *Grammatik Mac* (version 2.0) San Francisco, CA: Reference Software International.

Appendix 1: Readability Statistics for Individual Passages, Hokuriku University, 1994

Year/Test Code	94A1	94A2	94A3	94A4	94A5	94A6	94B1	94B2	94B3	94B4	94C1	94C2	94C3	94C4
Words	301	65	125	34	88	210	284	151	143	182	658	148	115	98
Syllables/Word	1.61	1.43	1.51	1.82	1.69	1.65	1.56	1.56	1.77	1.79	1.43	1.53	1.72	1.87
Words/Sentence	15.8	16.2	25	11.3	22	21	15.7	30.2	23.8	18.2	14.6	21.1	19.1	19.6
Flesch	55	69	54	41	41	46	59	44	33	37	71	56	42	29
Flesch-Kincaid	10	8	12	10	13	12	9	15	15	13	7	11	12	14
Fog	14	9	14	15	16	15	13	18	20	17	10	14	17	18

Year/Test Code	94D1	94D2	94D3	94D4	94E1	94E2	94E3	94E4	94F1	94F2	94F3	94F4	94Avg/passege
Words	320	146	147	127	300	116	95	78	535	136	95	124	185.4
Syllables/Word	1.7	1.56	1.69	1.66	1.44	1.7	1.8	1.58	1.37	1.56	1.7	1.55	1.58
Words/Sentence	24.6	36.5	29.4	31.7	11.1	14.5	19	19.5	10.7	15.1	19	20.6	18.02
Flesch	38	38	34	34	74	48	35	53	80	59	44	55	57
Flesch-Kincaid	14	17	16	17	6	10	13	11	5	9	12	11	10
Fog	17	20	20	20	8	14	17	14	7	12	16	13	13

Appendix 2a: Readability Statistics for Hokuriku University's Entrance Examinations, 1994

Year/Test Code	94A	94B	94C	94D	94E	94F	94-Avg.
Easy/Hard overall	—	—	—	Hardest	—	Easiest	—
Easy/Hard by Year	—	—	—	Hardest	—	Easiest	—
No. of Passages	6	4	4	4	4	4	4.33
Words	823	760	1019	740	589	890	803.5
Unique Words	441	351	456	405	330	400	397.16
Type-Token Ratio	53.58	46.18	44.75	54.73	56.03	44.94	50.035
Syllables/Word	1.61	1.65	1.52	1.66	1.57	1.46	1.58
Words/Sentence	18.2	19.4	16.1	28.4	13.3	12.7	18.02
Flesch	52	47	62	37	60	70	57
Flesch-Kincaid	11	12	9	15	8	7	10
Fog	14	16	12	19	11	9	13

Appendix 2b: Readability Statistics for Hokuriku University's Entrance Examinations, 1995

Year/Test Code	95A	95B	95C	95D	95E	95F	95-Avg.
Easy/Hard overall	—	—	—	—	—	—	—
Easy/Hard by Year	Easiest	—	Hardest	—	—	—	—
No. of Passages	3	3	3	3	3	3	3
Words	913	951	1021	883	666	778	868.66
Unique Words	418	455	432	441	311	367	404
Type-Token Ratio	45.78	47.84	42.31	49.94	46.7	47.17	46.62
Syllables/Word	1.39	1.42	1.48	1.51	1.45	1.45	1.45
Words/Sentence	16.9	16.3	21.2	15.2	15.1	22.2	17.81
Flesch	72	70	60	64	69	62	66
Flesch-Kincaid	8	8	10	8	8	10	8.66
Fog	10	11	14	12	10	13	11.66

Appendix 3a: Readability Statistics for Selected Universities, Inter-University Comparison

School	H.U. —	Kansai (private)	Hitotsu (public)	H.U. —	Keio (private)	Yokohama (public)
	Easiest	Easiest	Easiest	Hardest	Hardest	Hardest
No. of Passages	4	3	2	4	1	3
Words	890	1464	1243	740	986	858
Unique Words	400	239	295.5	405	515	159.33
Type-Token Ratio	44.94	48.98	47.55	54.73	52.23	55.71
Syllables/Word	1.46	1.46	1.43	1.66	1.65	1.64
Words/Sentence	12.7	18.98	16.47	28.4	19.04	25.71
Flesch	70	65.47	69.23	37	48.08	42.47
Flesch-Kincaid	7	8.4	7.7	15	11.28	13.61
Fog	9	9.92	10.18	19	13.26	15.91

Appendix 3b: Readability Statistics, Averages for Hokuriku, public, & Private Universities

	Center —	H.U. 94-Avg.	H.U. 95-Avg.	Public Avg.	Private Avg.
No. of Passages	3	4.33	3	3.4	2.3
Words	535	803.5	868.66	1135	1242
Unique Words	100.67	397.16	404	196.95	272.58
Type-Token Ratio	53.88	50.035	46.62	52.94	50.74
Syllables/Word	1.41	1.58	1.45	1.52	1.51
Words/Sentence	17.01	18.02	17.81	20.18	19.03
Flesch	70.35	57	66	58.29	60.4
Flesch-Kincaid	7.67	10	8.66	10.06	9.38
Fog	10.17	13	11.66	12.19	11.18