Miyagi University Research Journal

Trends in Productive, Receptive, and Multidimensional Vocabulary Metrics on University Entrance Exams in Northern Japan

Matthew Nall
Faculty of Foundational Academics, Miyagi University

【キーワード】
vocabulary assessment, open-ended assessment, close-ended assessment, lexical knowledge, standardized testing

【Correspondence】
Matthew Nall
宮城大学基盤教育群
nall@myu.ac.jp

【COI】
本論文に関して，開示すべき利益相反関連事項はない。

Received 2021.6.13
Accepted 2021.8.19

Abstract

Language assessment has played an integral part of the second language learning programs, both historically and globally. Since the 1980s, language assessment as an area of academic study has been gaining ground among academic circles related to language learning and teaching, and today it has become a thriving area of academic research. The present study examines English language assessments as they have been presented as part of the 2020 *zenki* university entrance exams at six public universities in northern Japan. The study identifies how these exams incorporate productive open-ended and receptive close-ended vocabulary assessments. These assessments serve as metrics used in order to evaluate university candidates before admission to the university. English entrance exam data in the present study was collected then analyzed by means of descriptive and inferential statistics. A chi-square for goodness of fit test was conducted to determine whether exam question categories differed from randomness as they appear on entrance exams in this region. Results show that there is little consensus among the six universities regarding how open- and close-ended vocabulary assessments are incorporated in the *zenki* entrance exams. Consequently, the universities do not seem to agree on how candidates should be evaluated concerning their receptive and productive vocabulary knowledge. Implications for the study include improved test validation methods, increased transparency by universities about entrance exams, and increased support for leaner autonomy and self-directed learning.

Introduction

Acquiring knowledge of the lexical features of language is one of the cornerstones in the effort to learn a new language (Nation, 2013; Zimmerman, 2009). For the second language (L2) learner, vocabulary knowledge is widely considered to be one of the most important aspects of language learning, as it is tied to the comprehension of content (Armbruster, 1992; Pearson et al., 2007), to reading (Anderson & Freebody, 1981) and listening comprehension, and to the production of language in written and oral forms. For L2 learners, having extensive vocabulary knowledge is important for those who are planning to use the language in social or academic contexts (Lightbown & Spada, 2013), and extensive vocabulary knowledge has been tied to overall academic success (Scott & Nagy, 1997). Vocabulary learning has remained one of the tenets of language learning programs both historically and internationally, but since the 1980s vocabulary acquisition and assessment have earned core positions in contemporary research circles concerned with language learning and teaching (Read, 2013).

In the field of second language acquisition (SLA), the increase in interest among researchers regarding second language vocabulary learning and teaching has prompted the development and use of various vocabulary assessments. Many of these vocabulary assessments are implemented as metrics which help identify the learner's overall vocabulary size in English as a second language. Here are some of the common vocabulary tests that are available:

- *The EFL Vocabulary Tests* (Maera, 1992)
- *The Productive Vocabulary Levels Test* (Laufer & Nation, 1999)
- *The Lex30* (Maera & Fitzpatrick, 2000)
- *The Computer Adaptive Test of Size and Strength* (Laufer & Goldstein, 2004)
- *The Vocabulary Size Test* (Nation & Beglar, 2007)
- *The Listening Vocabulary Levels Test* (McLean et. al, 2015)
- *The New Vocabulary Levels Test* (McLean & Kramer, 2015)
- *The Updated Vocabulary Levels Test* (Webb et. al, 2017)

As vocabulary assessments have gained in prominence in recent decades, the current research study considers English assessments as they are commonly presented as part of *zenki* university entrance exams at public universities in northern Japan. University entrance exams in Japan generally occur twice each year. The *zenki* exam is the first exam to take place, but unsuccessful applicants later have the opportunity to take the *kouki* entrance exam to try for admittance to the university.

Literature Review

As vocabulary knowledge has been viewed as a cornerstone in language learning, vocabulary assessments have maintained an important role in educational programs in various contexts. Such standardized tests may be designed with different purposes. These include: (a) diagnostic tests, (b) placement tests, (c) short and long-term achievement tests, and (d) proficiency tests (Nation, 2013). Questions that target vocabulary knowledge are important elements in each of these types of tests. Vocabulary assessment question types come in two basic forms—close-ended assessments, and open-ended assessments. These two methods of vocabulary assessment correspond to Nation's (2013) distinction between receptive and

Miyagi University Research Journal

productive language knowledge, respectively. Others have referred to this as the distinction between passive and active control of language over the complete spectrum of vocabulary knowledge and ability (Bravo & Cervetti, 2008). When utilized on tests in English as a foreign language (EFL) contexts, questions that target vocabulary knowledge can help to identify and evaluate the level of the test taker's receptive and productive vocabulary knowledge in the new language.

Historically, close-ended assessments have been one of the most common question types, and they are designed to measure the extent, or breadth, of the learner's receptive vocabulary knowledge (Resnick & Resnick, 1977). On tests, close-ended vocabulary assessments can be presented in a variety of ways. For example, multiple-choice type questions have perhaps been the most prominent form of close-ended assessment. These kinds of questions require that the test taker select the correct answer from a finite number of potential answers. Other close-ended question types include true/false questions, discrimination questions, check-list questions, as well as yes/no questions, among others. These kinds of questions require the test taker to draw upon their declarative knowledge of lexical forms in order to answer the questions correctly. That is to say that the test taker is not required to produce language in spoken or written form in order to answer close-ended questions.

In EFL contexts, close-ended assessments have also been a common form of language question on standardized tests. Their ubiquity in language testing has been due to their wide-spread familiarity and convenience, and they can be administered easily and quickly scored. There have been several prominent testing models that have influenced tests in this area. Classical Test Theory (Gyllstad et al., 2015) maintains that tests that evaluate a larger number of lexical items will have more valid and reliable results than shorter assessments. In contrast, Item Response Theory employs statistical models to extrapolate the actual test results, and estimate the breadth of the test taker's overall vocabulary knowledge. Vocabulary breadth indicates the number of words that the learner knows at least something about (Anderson & Freebody, 1981). This means that a shorter test can yield reliable and valid information concerning the overall breadth of the L2 learner's lexical knowledge. Item Response Theory has the benefit of allowing for shorter tests to be administered and for the reduction of test-taker fatigue while still producing valid and reliable results (Embertson, 1996).

In contrast, open-ended assessments are also important features of modern-day L2 language tests. Open-ended assessments "elicit productive knowledge by capturing how language learners use vocabulary in their talking and writing" (Bravo, 2018, p. 3069). While close-ended vocabulary assessments yield important information about the breadth of vocabulary knowledge (i.e., learners' receptive vocabularies), open-ended vocabulary assessments yield information about the depth of their vocabulary knowledge (i.e., how well they know a word), about how language learners are acquiring vocabulary, and about the test takers' ability to operationalize lexical knowledge in speaking or writing. Despite offering significant benefits in language assessment, some of the drawbacks of open-ended vocabulary assessments are that they are time consuming for the test-taker, and they are difficult to score.

Vocabulary Assessments in L2 Learning Contexts

In L2 learning contexts, test questions are inevitably multidimensional in nature. It is necessary to recognize that on tests designed to assess language learners on their lan-

guage acquisition progress, vocabulary items are inextricably involved in every kind of test question. This is the multidimensional nature of vocabulary assessments. Test questions are multidimensional because it is impossible for a test developer to design a test question that measures the test taker's knowledge of syntax, for example, without in some way also simultaneously testing vocabulary knowledge. Every kind of test question involves vocabulary in some way. Therefore, any kind of test question tests the learner's vocabulary knowledge in addition to whatever language feature or skill that the test developer is trying to target and evaluate with that particular test question.

In the modern globalized world, the assessment of productive language skills has become increasingly more important in L2 learning contexts. This phenomenon has occurred since the social turn in applied linguistics (Block, 2003) and the move in pedagogical practice toward incorporating more communicatively-focused language teaching models and approaches. Additionally, since the 1970s, advancements in understandings of cognitive science and psycholinguistics have changed the way that researchers approach problems related to learning vocabulary and second language acquisition. In turn, these advancements have influenced modern testing theories and models. Standardized tests have historically consisted of mostly close-ended questions. In recent decades, they have come to incorporate more open-ended questions in an attempt to adopt and standardize more contextualized and comprehensive assessment methods. These new assessments aim to identify not only the language learner's breadth of vocabulary knowledge, but also the depth of vocabulary knowledge and one's ability to produce language in either oral or written forms.

Open-ended questions offer some benefits in language assessment that close-ended questions do not. To say that a learner knows a word is much more complex than simply being able to define it (Johnson & Pearson, 1984; Nagy & Scott, 2000; Nation, 2013). Word knowledge consists of five factors of lexical knowledge. These features are acquired over time and experience, and include polysemy, interrelatedness, heterogeneity, incrementality, and multidimensionality (Nagy & Scott, 2000). Developing word knowledge is not only being able to define the word. This process is ongoing and takes place over long periods of time. It involves the acquisition of pragmatic ability and norms, as well as deep lexical and cultural knowledge. For the contemporary test developer, incorporating open-ended questions in language tests gives clues as to how the learner is acquiring language. Open-ended questions allow test developers to see language learners' "facility and flexibility with target vocabulary" (Bravo, 2018, p. 3076). As these aspects of language learning and vocabulary knowledge are increasingly important in today's globalized society, open-ended vocabulary assessments have established themselves as an important facet of L2 language tests in various contexts.

In the Japanese context, research concerning language assessment has been a thriving area of academic inquiry. However, to date, much of this research has been focused on national standardized tests, such as the *Senta Shiken* (i.e., Guest, 2008), which has been required for the majority of university applicants for decades. Additionally, in regard to entrance exams, many studies have focused on topics such as test washback (i.e., Brown, 2000; Mulvey, 1999), student motivation (i.e., Kikuchi, 2009), or the cognitive load of reading passages on these tests (i.e., Kikuchi, 2006). There has remained, however, a dearth of studies that investigate trends in university entrance exams with a focus on the types and nature of English vocabulary questions that are common, or on the distribution of such question types in specific geographical areas. The current study aims to fill this gap in understanding by examining the English question types presented in the 2020 *zenki* exams at

Miyagi University Research Journal

six prominent public universities in northern Japan.

Method

As open-ended vocabulary assessments have become more common features in language tests in EFL contexts, the current study offers a brief look into the ways in which various universities in northern Japan are currently incorporating open-ended questions in *zenki* entrance exams. The information presented in this study is offered with the aim of giving test developers a better idea about how their tests compare to those of other universities. Results of this study also have several important implications for L2 learners who are preparing to take university entrance exams in the near future.

In order to gather and analyze information about the use of different test question types commonly used in university entrance exams in northern Japan, the current research study draws upon the systematic nature of document analysis (Bowen, 2009), and analyzes the data by means of descriptive and inferential statistics (Creswell, 2008). Research questions 1 and 2 below are addressed using descriptive statistics. Research question 3 is addressed with inferential statistics, and a chi-square for goodness of fit test (Creswell, 2008) was conducted.

This research study analyzes and interprets the 2020 *zenki* university exams from six different public universities in northern Japan. The *zenki* university exam was chosen for this study because among the various kinds of university entrance exams, the *zenki* exam is prominent, and many successful applicants are granted admission to the desired university through this test. Every year, the *zenki* university entrance exam is administered in February, generally, and successful candidates are able to enter the university to start classes in April.

Research Questions

The research questions in the present study are as follows:

(1) What question types are commonly utilized in university entrance exams in northern Japan?

(2) To what extent are universities in this region utilizing open- and close-ended questions on university *zenki* entrance exams?

(3) Are vocabulary question categories on the *zenki* university entrance exams preferred equally among the six universities in this region?

Procedure

In order to gather the data necessary to answer the above research questions, the current researcher obtained the 2020 *zenki* entrance exams from six universities in northern Japan for analysis. Every year, after the actual *zenki* exams are administered, these exams are published and the exam contents become public knowledge. For the purposes of anonymity, the six universities will be referred to in this study as Universities A, B, C, D, E, and F. Each University designs its own *zenki* exam every year. The researcher devised the following coding scheme to categorize test question types. This coding scheme was based

on Bravo's (2018) description of open-ended questions, and close-ended question types are closely tied to Gyllstad et al.'s (2015) discussion of multiple-choice formats. In the present study, test questions were coded as follows: CJ = Close-ended Japanese question, OJ = Open-ended Japanese question, CE = Close-ended English question, and OE = Open-ended English question. In this coding scheme, CJ questions were considered to be questions presented on the test entirely in Japanese, and the test taker was required to select the correct answer from a finite number of possible answers, all of which were also in Japanese. OJ questions were open-ended in the sense that test takers had to produce written language in Japanese in some way. Some common OJ question types asked test takers to translate a phrase or sentence from English into Japanese, or to describe in Japanese what a certain word or phrase refers to in a text (i.e. *that*, *they*, or *these reasons*, etc.). CE questions required the test taker to select the appropriate English answer from a set of possible answers. These questions do not require that the test taker produce any language in writing. Finally, OE questions required the test taker to produce English in writing. This may require the test taker to produce English in the form of a single word or phrase, or to produce longer written text.

Having established the coding scheme for the present study, the researcher then reviewed each university's 2020 *zenki* entrance exam and systematically coded each test question accordingly. Detailed notes were kept on how each question was coded. Once the data collection was completed, the reviewer's notes were thoroughly reviewed, and the exams were re-checked by the researcher in order to make sure that test questions were not miscategorized. The data set was then entered into a spreadsheet for descriptive analysis, and IBM SPSS (version 26) for analysis by means of inferential statistics.

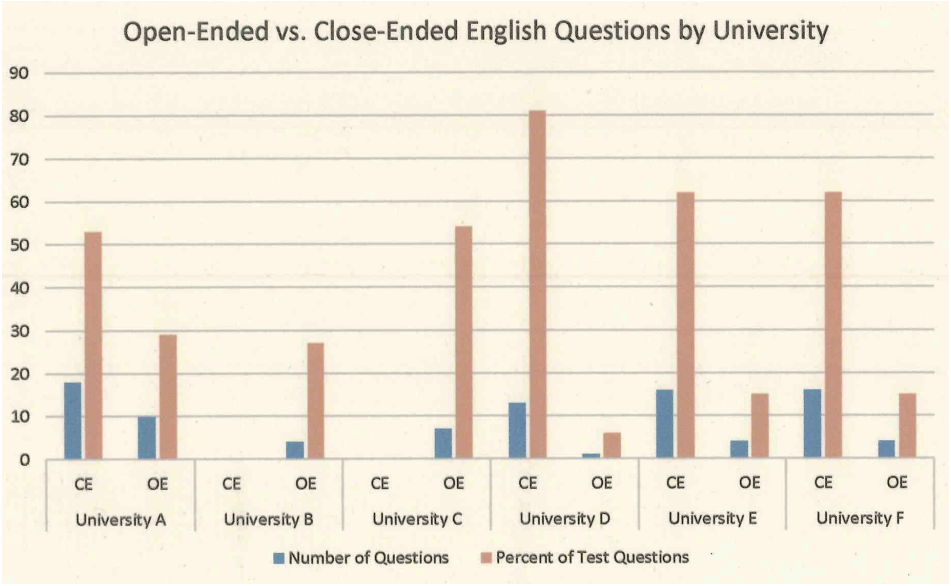
Results and Discussion

Firstly, in this study, none of the university entrance exams included CJ (close-ended Japanese) questions. For this reason, this category was excluded from the remainder of the data analysis, as the current researcher determined that this question category was of no consequence. The remaining data was organized in a spreadsheet, and bar graphs were generated in order to offer a visualization of the results. Figure 1 below displays each of the exams and their English question forms in bar graph format. Figure 2 shows the full range of test questions (OJ, CE, and OE) For each university, the blue bar displays the number of questions on each test in each question category. The orange bar shows the number of questions in each category as a percentage of the total number of questions on each exam. For example, in Figure 2, University A's exam included 6 questions that were open-ended Japanese questions (OJ) (i.e. the blue bar), which accounted for 18 % of the total test questions on that exam (i.e., the orange bar). On the same exam (i.e., University A), close-ended English questions made up the largest part of the test. This exam included 18 of these questions, which account for 53% of the test questions. Finally, on University A's exam, there were 10 open-ended English questions that accounted for 29% of all of the test questions.

Miyagi University Research Journal

Figure 1

Descriptive Analysis of Exam Question Types by University



Note: OJ = open-ended Japanese question, CE = close-ended English question, and OE = open-ended English question.

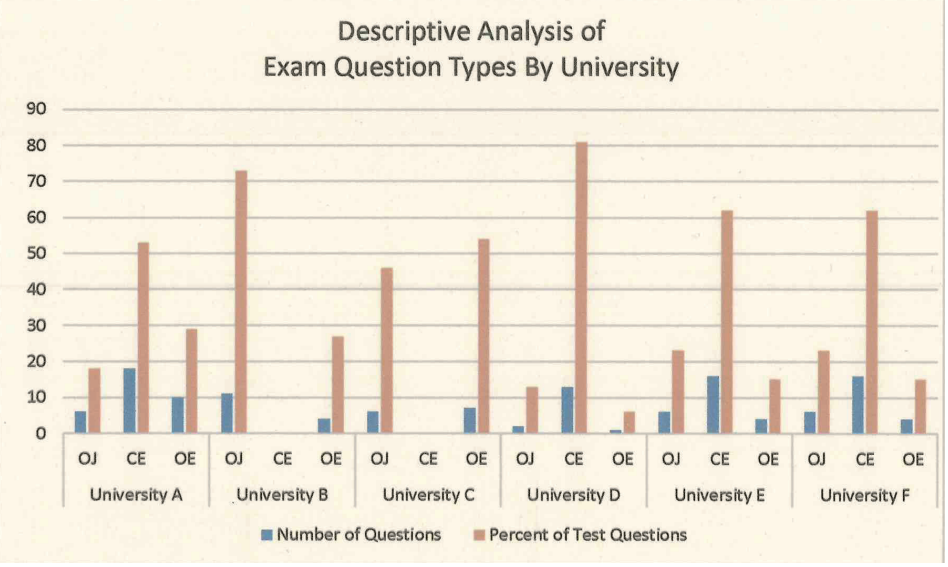
To address research question 1, Figure 1 gives some important revelations about how the different kinds of test questions are being utilized on entrance exams. There is wide variation among different universities' assessment strategies, as there appears to be little consensus as to how these categories are incorporated into the exams. For example, Universities B and C's exams included zero close-ended English questions. In contrast to this, University D had the highest occurrence of close-ended English questions, with this category making up over 80% of the total test questions.

Regarding research question 2, Figure 2 offers important clues about how open- and close-ended question types are distributed on the exams. As open-ended assessments have become increasingly important features of standardized tests, Figure 2 displays the open-ended English questions for universities A through F. Once again, it can be clearly seen that there is little agreement as to how these types of questions are incorporated into exams. University C has the highest percentage of open-ended English questions, which made up over 50% of the total exam questions. University D, on the other hand, put little emphasis on open-ended English questions, as this category only made up 6% of the exam questions. Concerning both open- and close-ended exam questions that required the test taker to answer in English, it is also interesting to take note of the proportions of open- and close-ended questions within each exam. This data is presented in Figure 2. While some exams had a larger percentage of test questions that were close-ended (i.e. University A, D, E, and F), others showed a preference for a larger proportion of test questions to be open-ended (University B and C). This again shows a lack of standardization or consensus among the different universities regarding the use and frequency of the different question type categories.

Miyagi University Research Journal

Figure 2

Open-Ended vs. Close-Ended English Questions by University



Finally, to address research question 3, IBM SPSS (version 26) was utilized to determine whether test question types are preferred equally by universities in this region. For this, the frequency of the question types (OJ, CE, and OE) were calculated for all of the 6 entrance exams together (OJ = 37; CE = 63; OE = 20; Total questions = 120). A chi-square goodness of fit test was used to test whether the pattern of test question types used on university entrance exams in northern Japan differed from randomness (i.e. they showed a statistically significant preference for one or more question types). Expected frequencies in all cells were greater than five. Results showed a statistically significant difference between CE questions (63), and OJ questions (37) or OE questions (20), $\chi^2(2, n=120) = 23.45$, $p = .001$. This indicates that the question types are not distributed equally, and that overall, universities in this region had a statistically significant preference for close-ended English questions in the entrance exams.

Discussion and Conclusion

Concerning research questions 1 and 2, the primary conclusion from the descriptive data analysis is that there is little consensus among the six universities about how and to what extent open- and close-ended vocabulary questions are incorporated in the exams, and about how the candidates' vocabulary knowledge and abilities in these areas are tested and evaluated. The six universities evaluate the candidates' productive and receptive vocabulary knowledge in very different ways, often providing results that do not adequately display the full range of the test taker's receptive and productive vocabulary knowledge and abilities. For example, some universities (like Universities D, E and F in the present study) administer an entrance exam with a heavier weight going to close-ended English vocabulary questions. This means that these universities will be unable to appropriately judge the candidates' depth of English vocabulary knowledge and their language production abilities. In contrast, exams (like those of Universities B and C) that focus more exclusively on open-ended vocabulary question categories are likely to be inept tools for evaluating the candidates' overall breadth of vocabulary knowledge, as receptive L2 vocabularies tend to be much larger than productive vocabularies (Nation, 2013). These differences among the

six universities' entrance exams are significant. As the individual exams put more or less emphasis on productive or receptive vocabulary knowledge by means of favoring either open- or close-ended questions, the candidates' vocabulary knowledge is therefore evaluated in very different manners, painting significantly different pictures of the candidates' overall vocabulary knowledge and abilities. The graphs presented in Figure 1 and Figure 2 display these inequalities in how candidates are tested and evaluated. The data provided in the present study offer a reference for exam developers, so that they can quickly understand how their entrance exam compares to the exams of other universities. This information also shows that there is a tendency for universities in this region to assign a heavier percentage of exam questions to close-ended question types (i.e. University A, D, E, and F). This may be due either to the familiarity and convenience of these types of questions, or the ease of grading such questions. Some universities in the data set, however, put exclusive focus on open-ended English questions (University B and C). These tests are focused on identifying and evaluating test takers' productive knowledge of the language, and their ability to operationalize lexical items and display the depth of their vocabulary knowledge.

Concerning research question 3, the chi-square for goodness of fit test also confirmed the above results. The p value in this test was less than .05, indicating that there was a statistically significant difference among the frequency of test question types used in entrance exams among the six universities. The question types are not distributed equally, and overall, there is a strong preference for close-ended questions that target only the learner's receptive vocabulary knowledge. This can be seen as being detrimental because productive vocabulary knowledge is becoming increasingly important in the globalized world, but it accounts for a smaller portion of test questions on some university entrance exams. From the perspective of the learner, this study may offer some guidance about how students might more appropriately study for university entrance exams. Being informed by data concerning the distribution of question types and their frequencies on entrance exams, students are able to guide their own learning in an appropriate manner.

Overall, the results of this study hint that test reliability and validation need to be given more consideration during the test development process (Schmitt et. al, 2020). This is made evident by the lack of standardization or agreement among the various universities regarding how and to what extent open- and close-ended vocabulary assessments are incorporated into the university entrance exams, and by observing the differences among universities concerning how receptive and productive vocabulary knowledge are tested and evaluated. In this regard, "Test developers need to be very clear WHY they are making a test (purpose), WHO it is intended for and in what CONTEXT, and WHAT ASPECT(S) OF VOCABULARY KNOWLEDGE they are trying to measure" (Schmitt et. al, 2020, p. 111) [emphasis is included in the original source]. Test reliability and validity can be improved through the test validation process. Concerning test validation, open-ended assessments pose significant challenges as they are difficult to score fairly and consistently. This area deserves more attention from the various universities during the test development and test validation processes. Rubrics are effective tools in language assessment, and they can also be utilized for learning purposes as well (Elsheikh, 2018). Future studies on university entrance exams may examine or develop and evaluate different rubrics in order to help exam graders to fairly and consistently evaluate learners' productive vocabulary knowledge on open-ended vocabulary assessments.

The current study has some limitations. Firstly, this study only analyzes exam data from six public universities in a specific region of northern Japan. The validity and reliability

of results could be improved if future studies included both public and private university entrance exam data, and if the scope of the research were expanded to include universities at a national level. It may also be of interest to conduct future studies with a historical or longitudinal scope. This type of study would offer clues as to how entrance exams have evolved over time concerning their utilization of the different question categories, and about how universities are evaluating candidates in terms of their productive and receptive vocabulary knowledge. This would also offer clues as to how language learning policy is taking place or changing over time in Japan regarding receptive versus productive skills development and how these aspects of language learning are evaluated on standardized exams. The results of such studies would likely give direction for the development of future exams as well as inform future educational policy.

Finally, the most impactful limitation of the current study is that the actual points allotted for each exam question could not be used in the data collection or analysis, as this information is not public knowledge. If this were made public knowledge, future studies would offer a more appropriate picture of the actual weight of the different question categories on the entrance exams, and it would give candidates a better idea about how they can direct their vocabulary study efforts more effectively. This knowledge would have several important effects from the learning perspective. It would help candidates to become more autonomous learners, to have increased metacognitive awareness, and to be able to efficiently and effectively employ self-directed learning as they prepare to take university entrance exams, all which have positive impacts for the learner. With this in mind, another implication of the present study is for universities to allow for more transparency regarding their entrance exams and how they are scored and evaluated. Currently, the test questions themselves are made publicly available after the administration of the actual entrance exams, but if the points allotted to each question were also made available, then future test takers would have increased control over their own learning.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Gutherie (Ed.), *Comprehension and teaching: Research reviews*. International Reading Association.

Armbruster, B. B. (1992). Vocabulary in content area lessons. *The Reading Teacher*, 45, 550-551.

Block, D. (2003). *The social turn in second language acquisition*. Edinburgh University Press.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40.

Bravo, M. A. (2018). Open-ended vocabulary assessments. In J.I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (1st Edition, pp. 3069-3077). John Wiley & Sons.

Bravo, M. A., & Cervetti, G. N. (2008). Teaching vocabulary through text and experience in content areas. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 130-149). International Reading Association.

Brown, J. D. (2000). University entrance examinations: Strategies for creating positive washback on English language teaching in Japan. *Shiken 3*(2), 2-7.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education, Inc.

Elsheikh, A. (2018). Rubrics. In J.I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (1st Edition, pp. 5199-5207). John Wiley & Sons.

Embertson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341-349.

Guest, M. (2008). A comparative analysis of the Japanese university entrance *Senta Shiken* based on a 25-year gap. *JALT Journal*, 30(1), 85-104.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats. *ITL - International Journal of Applied Linguistics, 166*(2), 278-306.

Miyagi University Research Journal

Johnson, D. D., & Pearson, P. D. (1984). *Teaching reading vocabulary*. Rinehart and Winston.

Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal, 28*(1), 77-96.

Kikuchi, K. (2009). Listening to our learners' voices: What demotivates Japanese high school students? *Language Teaching Research, 13*(4), 453-471.

Laufer, B., & Goldenstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 36-55.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.

McLean, S., & Kramer, B. (2015). The creation of a New Vocabulary Levels Test. *Shiken, 19*(2), 1-11.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19* (6), 741-760.

Maera, P. (1992). *EFL vocabulary tests*. University College, Swansea: Centre for Applied Linguistics. https://www.lognostics.co.uk/vlibrary/meara1992z.pdf

Maera, P., & Fitzpatrick, T. (2000). Lex 30: An improved method of assessing productive vocabulary in an L2. *System*, 28, 19-30.

Mulvey, B. (1999). A myth of influence: Japanese university entrance exams and their effect on junior and senior high school reading pedagogy. *JALT Journal, 21*(1), 125-142.

Nagy, W. E., & Scott, J. A. (2000). Vocabulary processing. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, 269-274.

Nation, I. S. P. (2013). *Learning vocabulary in another language (2nd ed.)*. Cambridge University Press.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282-296.

Read, J. (2013). Second language vocabulary assessment. *Language Teaching, 46*(1), 41-52.

Resnick, D. P., & Resnick, L. (1977). The nature of literacy: An historical exploration. *Harvard Educational Review, 47*, 370-385.

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching*. Cambridge University Press.

Scott, J. A., & Nagy, W. (1997). Understanding the definitions of unfamiliar verbs. *Reading Research Quarterly*, 32(2), 184-200.

Webb, S., Sasao, Y., & Balance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL – International Journal of Applied Linguistics, 168*(1), 34-70.

Zimmerman, C. B. (2009). *Word knowledge: A vocabulary teacher's handbook*. Oxford University Press.