

## Using quantitative research approaches to place students in Japanese university English language classes

**Kristy King Takagi**

University of Fukui, Japan

[kjktakagi@hotmail.com](mailto:kjktakagi@hotmail.com)

### **Abstract**

In university English language programs in Japan, students are often tested soon after admission to the university in order to determine their level of English proficiency, so that they can be placed into classes at advanced, intermediate, and basic levels, a result that is generally seen as advantageous for both students and teachers. However, how these test scores are to be used to create placement lists is not always given extensive consideration, especially from a statistical perspective. For example, programs might simply combine test scores or use them in another relatively simple way, probably in part because the specialties of program administrators and teachers tend to be in areas such as linguistics and language, rather than quantitative research methods. In this paper, I will first explain the placement testing situation in Japan and briefly discuss tests frequently used, and then explain how placement committees might improve their placement process by using one of a number of ideas and approaches from quantitative research methods, such as  $z$  scores,  $T$  scores, factor analysis, and Rasch analysis, to create English class placement lists of students.

### **Key words**

Quantitative research approaches, English language class placement testing

### **Introduction**

Whereas university entrance examinations in Japan have been the object of study (and much criticism) for many years, placement tests in Japan, and elsewhere around the world, have received little attention. According to Wall, Clapham, and Alderson (1994) there has been little research on the “nature” of placement tests, and much more study is needed on appropriate ways to validate these commonly used tests (p. 321). Culligan and Gorsuch (1999) also said that there has been “surprisingly little research” on commercial tests for EFL (English as a Foreign Language) use in Japan, as well as few accounts of how tests can be developed for in-house use (p. 8).

When English language placement test use in Japanese universities was examined in the few studies that are available, various problems were found. Wistner and Sakai (2008) said that validity evidence for placement test use was not being gathered, tests had not been examined adequately for reliability, and students were often placed in classes on the basis of such unexamined test scores. As a result, they said that chances can often be high that students are “arbitrarily placed” in English classes in Japanese universities, and that the proficiency levels can be quite diverse in one class (p. 1054). These mixed-level classes cause obvious difficulties for teachers, but Wistner and Sakai also noted that incorrect class placement can “adversely affect” students (p. 1054). Besides these problems, Shimizu (2000) said that “there is no established precedence or theoretical foundation for test development nor a guideline for statistical analysis that each school should follow” (p. 243). Although statistical analysis of tests includes much more, the focus of this paper will be statistical analysis that can be performed on test scores in order to place students into English classes, in particular, creation of  $z$  scores and  $T$  scores, factor analysis, and Rasch analysis. Before these statistical methods are discussed, a brief description of tests used in Japanese university English programs will be presented.

### **Tests used for English language class placement testing in Japan**

Despite clear advantages to English class placement, some university programs in Japan do not place students by proficiency level in their English classes. According to Shimizu (2000), who surveyed teachers at 200 Japanese universities, one reason cited for not using placement testing was concern that students at the basic level of study would suffer psychologically from being placed in that class level; however, most objections seemed to focus on problems with logistics and with the increased work load that would accompany testing. Even so, it appears that English class placement testing is being increasingly used in Japan. Shimizu cited results of another researcher’s 1983 survey in which approximately 4% of those who responded said that their universities conducted English class placement testing, while approximately 32% of those who responded to Shimizu’s 2000 survey did so.

When Japanese university English programs do administer placement tests to sort students into classes at different levels of proficiency, they use either a standardized test or a test developed in house. A variety of standardized tests are used, such as TOEFL (Test of English as a Foreign Language), TOEIC (Test of English for International Communication), MEPT (the Michigan English Placement Test), OPT (the Oxford Placement Test), SLEP (the Secondary Level English Proficiency Test), QPT (the Quick Placement Test), the CEFR interview (Common European Framework of Reference for

Languages), CELT (Comprehensive English Language Test), STEP or *Eiken* tests (the Test in Practical English Proficiency), JACET (Japanese Association of College English Teachers) tests, tests related to course textbooks, and others.

None of the standardized tests studied was wholly endorsed. As Wistner, Sakai, and Abe (2009) pointed out, TOEFL is the most researched test in applied linguistics; however, it is usually not considered to be appropriate for placement testing. As Brown and Hudson (1998) noted, its focus is too wide and general; results are generally not precise enough for placing students into classes in a particular program, nor are results directly related to the courses taught there. In addition, TOEFL results do not include individual item information, therefore precluding the item analysis that is necessary to determine whether the test is appropriate in a program or not. Similar concerns apply to the TOEIC, as well. Though there are no reports concerning how well either test has worked as a placement test in Japanese universities, it is reasonable to assume that they have not worked well. As for MEPT, Wistner, Sakai, and Abe (2009) found that the test subscores were low in reliability and did not effectively divide students into two groups; in similar fashion, Brooke, Aden, Al-Kuwari, Christopher, Ibrahim, Johnson, and Souyah (2012) concluded that the MEPT reading passages were too short to allow proper evaluation of a variety of reading skills. As for the OPT, Wistner et al. (2009) found that the test was problematic in that the reliability of the listening subtest was low. Brooke et al. (2012) also criticized the listening section and deemed it invalid for their program, as it did not correlate with other English measures. The SLEP was judged by Culligan and Gorsuch (1999) to be invalid for their program due to measurement error, failure to discriminate between high and low scorers, and lack of a speaking section. The QPT also did not discriminate well between high and low scorers, and more than half of the test items were too difficult or too easy for the students (Westrick, 2005). Finally, the CEFR, despite its growing popularity, has also been criticized as well, particularly in terms of its theoretical underpinnings (Alderson, 2007; Hulstijn, 2009).

In short, the little research that is available regarding standardized tests used for English class placement in Japanese universities reveals that none of the tests studied is an ideal fit to any program. Wistner and Sakai (2008) simply concluded that standardized tests “do not work well” for placement in English programs in Japanese universities (p.1047). Instead of standardized tests, they advocated creating in-house tests. Many other researchers concur, but, as Culligan and Gorsuch (1999) cautioned, developing in-house tests requires testing expertise and teamwork, is an “arduous” process that requires ongoing evaluation and revision, and can have disappointing results (p. 9). In fact, Wistner and Sakai (2008) found that their own in-house test performed poorly. It was too easy for students, had high error estimates, and did not

divide students into two groups. Despite their own obvious testing expertise, the test they created was not successful.

Clearly, the type of placement test to be used in a university English program in Japan is one issue that needs further study. According to Shimizu (2000), in her survey of universities across Japan, many teachers hope for some sort of standardized, national placement test, and are also concerned about the administrative burden that placement testing can bring to already overworked teachers.

Indeed, placement testing entails many tasks and requires expertise, as well. Test evaluation, revision, and validation require knowledge of classical test theory and item response theory, particularly Rasch Analysis. Though in-depth discussion of these areas or any other statistical matter is beyond the scope of this short paper, I will generally discuss the advantages of using  $z$  scores,  $T$  scores, factor analysis, and Rasch analysis to analyze student test scores, and then place students into proficiency levels for class placement.

After students sit the English class placement test, placement team members will have test scores in hand. If only one test score is to be used, creating the placement list is not terribly difficult. Raw scores can simply be arranged in order and students can be separated into classes on the basis of these scores. Administrators of some programs simply go down the list of scores and mark off the list into relatively equal class groupings. If the placement team wanted to divide the group of students in a more meaningful way, such as into advanced, intermediate, and basic classes, one method they could use would employ the mean, or average score, and the standard deviation. Students whose scores were one standard deviation above the mean or higher could be designated as advanced; intermediate students would have scores in the middle, and the basic group would have scores one standard deviation below the mean or lower. For example, if students took a 100-point placement test, and the mean was 60, and standard deviation 20, placement scores for the advanced, intermediate, and basic groups would be: 80 and above; 41 to 79; and 40 and below, respectively. If the numbers of students in the advanced and basic groups were very small, this division using one unit of standard deviation might not be practical; however, some use of standard deviation would make the class labels more meaningful. There are a number of decisions that must be made, but working with only one test score is relatively straightforward.

If a placement test included two or more scores, the placement team would have more decisions to make and more tasks to complete. If the team wanted to use or weight the two scores equally, but the scores were not on the same scale, they could not simply add the scores. For example, if the total possible score of one test was 30

points, and the possible score for the second test was 100 points, the team would need to create standardized scores for each. Afterwards, the two standardized scores could be averaged. Brown (1996) said that language teachers often find standardized scores “somewhat mysterious” (p. 133); however, they are not difficult to comprehend. A standardized score makes clear how far a student’s score is away from the mean, in units of standard deviation. The *z* score and the *T* score are two types of standardized scores which placement testing teams can make use of. To determine the *z* score, one must subtract the mean from the student’s score, and then divide that number by the standard deviation (p. 133). For example, using the test scores above (of a 100-point test with a mean of 60 and standard deviation of 20), the *z* score for a student who scored 80 would be  $80-60/20$ , or +1.0. If a second student scored 60, that student’s *z* score would be  $60-60/20$ , or 0. (That is, the mean of *z* scores is 0.) Finally, if a student scored, 40, his *z* score would be  $40-60/20$ , or -1.0. Because many people are not comfortable thinking about scores in such terms, testing officials often use a standardized score called a *T* score. *T* scores are calculated by multiplying the *z* score by 10, and then adding 50 (p. 135). Therefore, a student with a *z* score of +1.0 has a *T* score of 60; a student with a *z* score of 0 has a *T* score of 50; and a student with a *z* score of -1.0 has a *T* score of 40. Though standardized, these *T* scores look like actual scores and are more easily understood by people who have limited experience with testing. Therefore, after the placement team creates *T* scores by using both the raw and *z* scores, they can create class placement lists based on the *T* scores, and then more easily communicate the results.

Another interesting method for using a number of different scores to create student placement lists for English classes at Japanese universities was suggested by Everitt and Hothorn (2011)—principal components analysis, or PCA. They explained that “applying principal components to the observed examination results and using the students’ scores on the first principal components” would “provide a measure of examination success that maximally discriminates between them (p. 62). Although PCA requires a somewhat higher level of statistical expertise than that required to create and use *z* or *T* scores, teachers and administrators who complete a bit of study and practice could successfully use it. As Field (2005) explained, PCA is a less complex form of factor analysis, and in his guide, explains how to complete such an analysis using SPSS (Statistical Package for the Social Sciences). Another useful guide that explains this analysis in a straightforward and clear manner is the well-known guide to using SPSS with Windows and Macintosh, by Green and Salkind (2011).

Finally, Rasch analysis provides yet another interesting and valuable method for creating English class placement lists from test scores; it also offers a wealth of important information for placement testing teams. In their well-known text,

*Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Bond and Fox (2007) provide an in-depth introduction to Rasch analysis. In addition, a number of researchers have used Rasch analysis to study placement tests, and their work can inform the efforts of placement testing teams. For example, Gorsuch and Culligan (2000) explained how they used Rasch analysis in making English class placement decisions. They explained that Rasch analysis is especially useful for placement since it simultaneously provides estimates of student ability and test item difficulty, and it creates a model that, though based on the original data, is “thought to hold for all students who take the test in the future” (p. 318). Besides student ability and item difficulty measures, this analysis also provides information about group separation, bias against particular test takers, reliability, fit, and much more. In terms of creating placement lists, the testing team could simply use ability estimates derived from the analysis to arrange students in order. Or they could investigate cut scores, as well, as Gorsuch and Culligan did. In their case, they matched Rasch ability estimates to cut scores established from raw scores, and then compared the two approaches, and the relative distribution of students for each method. By studying cut scores further, placement teams could learn more about the cut scores that would be most appropriate for their own students and English classes.

### **Conclusion**

In conclusion, there is little research available on English placement tests in Japan and around the world, but what little there is suggests that the contention of Wall, Clapham, and Alderson (1994) remains true; there is still insufficient research on the “nature and validation” of placement tests (p. 321). Part of the research that is needed is related to the process discussed in this paper: how to choose the best English placement tests, and then properly analyze student test scores, in order to correctly evaluate students’ proficiency levels, and then place them in the most appropriate English language classes. Although this part of the statistical analysis may appear small, the work involved would allow placement teams the opportunity to study test validity as they ask for feedback from teachers and students regarding the appropriateness of placement decisions; compare test scores with other similar measures; and carefully study items and sections of the placement test. As placement teams work through this more basic part of the placement process in an informed, patient, and persistent manner, knowledge of both the nature and validation of placement tests will gradually but surely grow.

### **References**

- ALDERSON, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91(4), 659-663.
- BOND, T. G., & FOX, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, N. J.: Lawrence Erlbaum.
- BROOKE, K., ADEN, M., AL-KUWARI, N., CHRISTOPHER, V., IBRAHIM, M., JOHNSON, B., & SOUYAH, O. (2012). Placement testing in an EFL context. *TESOL Arabia Perspectives*, 19(2), 13-20.
- BROWN, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, 23(1), 65-83.
- BROWN, J. D. (1996a). English language entrance examinations in Japan: Problems and solutions. *Proceedings of the JALT 1995 conference*, Nagoya, Japan, 272-283.
- BROWN, J. D. (1996b). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
- BROWN, J. D., & HUDSON, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- CULLIGAN, B., & GORSUCH, G. (1999). Using a commercially produced proficiency test in a one-year EFL curriculum in Japan for placement purposes. *JALT Journal*, 21(1), 7-28.
- EVERITT, B. & HOTHORN, T. (2011). *An introduction to applied multivariate analysis with R*. New York: Springer.
- FIELD, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.
- FULCHER, G. (1997). An English language placement test: Issues in reliability and validity. *Language Testing*, 14(2), 113-138.
- GORSUCH, G. J. & CULLIGAN, B. (2000). Using item response theory to refine placement decisions. *JALT Journal*, 22(2), 315-325.
- GREEN, S. B. & SALKIND, N. J. (2011). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, N. J.: Pearson Prentice Hall.
- HULSTIJN, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4), 663-667.
- MURPHEY, T. (2001). Nonmeritorious features of the entrance exam system in Japan. *The Language Teacher*, 25(10), 1-4.
- SHIMIZU, Y. (2000). Survey research on the use of placement tests at four-year universities in Japan. Retrieved June 6, 2015, at [http://www.ritsumei.ac.jp/acd/re/k-rsc/lcs/kiyou/14-1/RitsIILCS\\_14.1pp.231-243SHIMIZU.pdf](http://www.ritsumei.ac.jp/acd/re/k-rsc/lcs/kiyou/14-1/RitsIILCS_14.1pp.231-243SHIMIZU.pdf)

- WALL, D., CLAPHAM, C., & ALDERSON, J. C. (1994). Evaluating a placement test. *Language Testing*, 11(3), 321-344.
- WESTRICK, P. A. (2005). Score reliability and placement testing. *JALT Journal*, 27(1), 71-93.
- WISTNER, B., & SAKAI, H. (2008). Rasch analyses of English language placement tests. In Bradford Watts, K., Muller, T. & Swanson, M. (Eds.), *JALT2007 conference proceedings* (pp.1045-1055). Tokyo: Japan Association for Language Learning.
- WISTNER, B., SAKAI, H., & ABE, M. (2009). An analysis of the Oxford Placement Test and the Michigan English Placement Test as L2 proficiency tests. Retrieved June 8, 2015 from <https://www.hosei.ac.jp/bungaku/museum/html/kiyo/58/articles/Wistner.pdf>

### Contact

Kristy King Takagi  
University of Fukui  
3-9-1 Bunkyo  
Fukui City  
Fukui 910-8507  
Japan  
[kjktakagi@hotmail.com](mailto:kjktakagi@hotmail.com)