タイトル	Assessing Graphophonemic Awareness in an English Classroom in Japan
作成者（著者）	市山, 陽子
公開者	東邦大学
発行日	2018.02
ISSN	03877566
掲載情報	東邦大学教養紀要. 49. p.33-40.
資料種別	紀要論文
著者版フラグ	publisher
JaLCDOI	info:doi/10.14994/toho.liberal.arts.rev.49.33
メタデータのＵＲＬ	https://mylibrary.toho-u.ac.jp/webopac/TD28004342

Toho Liberal Arts Review No. 49 (2017)
東邦大学教養紀要　第49号（2017）

Assessing Graphophonemic Awareness in an English Classroom in Japan

Yoko Ichiyama

Abstract

Despite the importance of assessing graphophonemic awareness in foreign-language education, especially of English as a foreign language (EFL), little attention has been paid to such assessment at tertiary-level institutions in Japan. On the contrary, Japanese universities tend to focus either on developing students' general English communication skills or on teaching students English for specific purposes (e.g., business English). To address that gap in practice, this paper describes the process of assessing graphophonemic awareness in an EFL classroom.

Introduction

The development of graphophonemic awareness—that is, the ability to recognize relationships between graphemes and phonemes—is crucial not only in first-language acquisition (L1) but also in second-language (L2) and foreign-language (FL) learning [1-4]. Studies have shown that readers given explicit instructions regarding grapheme-phoneme relationships become more skilled in reading than readers who do not receive such instructions [5]. Despite the importance of cultivating graphophonemic awareness in language learning, even in later stages of learning [6, 7], relatively little attention has been paid to developing such awareness among English-language learners in Japan [8]. As [8] has pointed out, acquiring English graphophonemics is more difficult for Japanese EFL learners given the different orthographic depths of two languages. Japanese language systems, especially hiragana, katakana, and romaji, have shallow orthographic depths in which letter-sound relationships are regular [9]. However, English exhibits deep orthographic depths containing no one-to-one relationships between graphemes and phonemes.

The 2012 version of the Course of Study, a national syllabus proposed by Japan's Ministry of Education, Culture, Sports, Science, and Techhnology (MEXT) moved the introduction of romaji, a foreign style of writing Japanese that uses an alphabet [9], from the fourth year to the third year of primary school [10]. As Table 1 shows, each hiragana or katakana character can be transcribed into romaji based on its sound [11]. Romaji, especially modern romaji, refers to characters using the Hepburn system of romanization developed by J. C. Hepburn in the nineteenth century.

Table 1 *Japanese Hiragana, Katakana, and Romaji with International Phonetic Alphabet* (*Suski, 1931*)

H	K	R	I	H	K	R	I	H	K	R	I	H	K	R	I	H	K	R	I
あ	ア	a	[a]	い	イ	i	[u]	う	ウ	u	[ɯ]	え	エ	e	[e]	お	オ	o	[o]
か	カ	ka	[ka]	き	キ	ki	[ki]	く	ク	ku	[kɯ]	け	ケ	ke	[ke]	こ	コ	ko	[ko]
さ	サ	sa	[sa]	し	シ	shi	[ɕi]	す	ス	su	[su]	せ	セ	se	[se]	そ	ソ	so	[so]
た	タ	ta	[ta]	ち	チ	chi	[tɕi]	つ	ツ	tsu	[tsɯ]	て	テ	te	[te]	と	ト	to	[to]
な	ナ	na	[na]	に	ニ	ni	[ni]	ぬ	ヌ	nu	[nɯ]	ね	ネ	ne	[ne]	の	ノ	no	[no]
は	ハ	ha	[ha]	ひ	ヒ	hi	[çi]	ふ	フ	fu	[ɸu]	へ	ヘ	he	[he]	ほ	ホ	ho	[ho]
ま	マ	ma	[ma]	み	ミ	mi	[mi]	む	ム	mu	[mɯ]	め	メ	me	[me]	も	モ	mo	[mo]
や	ヤ	ya	[ja]					ゆ	ユ	yu	[jɯ]					よ	ヨ	yo	[jo]
ら	ラ	ra	[ɾa]	り	リ	ri	[ɾi]	る	ル	ru	[ɾɯ]	れ	レ		[ɾe]	ろ	ロ		[ɾo]
わ	ワ	wa	[wa]	ゐ		wi	[wi]					ゑ		we	[we]	を	ヲ		[wo]
ん	ン	n	[n]																

In Japan, since the introduction of English reading education, especially of the English alphabet, continues to occur in junior high school, Japanese students learn to read in the romaji alphabet before learning to read in the English alphabet. However, that trend tends to confuse Japanese learners of English, for whereas the romaji alphabet has one-to-one relationships between graphemes and phonemes, English does not [11, 12].

To cultivate students' graphophonemic awareness, measuring their level of such awareness at earlier stages of language programs is crucial to determining what sort of support offered in language classrooms would be beneficial [12, 13]. At the same time, validating test items is important to review whether they can offer insights into the abilities and knowledge of test takers that can adequately guide the development of future test items [14]. Using Rasch analysis, which provides log odds ratios of probability used to transform nominal scales between individual items and test takers into interval scales, is recommended for validating the efficiency and efficacy of tests [15–17]. Since Rasch analysis requires relatively small samples of item and test takers, it is moreover ideal for performing validations in settings with fewer than 100 test takers [18–20].

The mentioned trends in English-language education in Japanese universities, the need to conduct assessments that offer useful insights, and the benefits of conducting Rasch analysis encouraged the use of such analysis to validate 54 test items assessing the graphophonemic awareness of university students at a tertiary-level institution in Japan.

Methods

A graphophonologic awareness test consisted with 54 items was administered as a part of a final test at a tertiary-level institution in Japan in 2015. Each item consisted of two English words with an underlined grapheme, as shown in the following example:

If the sounds of the underlined letters of both words are the same, then write "Y", if different, then write "X" instead. If you do not know the sound of the grapheme on the left, then write "L"; if you do not know the sound of the grapheme on the right, then write "R"; and if you do not know the sounds of both graphemes, then write "B" instead.

1. ball same ______

In short, test takers needed to identify whether phonemes of the underlined graphemes were the same or not.

A set of 108 grapheme–phoneme relationships informed the material to which test takers responded during the study. Of the 108 graphemes, 39 graphemes (36.2%) assessed test takers' ability to identify a grapheme with a phoneme that does not exist in Japanese pronunciation in the final test; 48 graphemes (44.4%) required test takers to identify a grapheme pronounced in a phoneme that exists in Japanese but that does not follow a one-to-one grapheme-phoneme relationship in Japanese; and the remaining 21 graphemes (19.4%) targeted the test takers' ability to identify a grapheme with a phoneme that follows a one-to-one grapheme–phoneme relationships in Japanese. Of the 54 test items, 11 items (20.3%) aimed to assess English consonants, whereas the 43 others aimed to assess English vowels.

The computer program Winsteps, Rasch version 3.81.0, was used to analyze the data [18]. The test was analyzed for its separation and reliability, targeting, item fit, and unidimensionality. [20] has observed that the Winsteps manual states that any person separation index less than 2.0 with a person reliability coefficient less than 0.8 indicates test item's inability to divide test takers by performance and that additional items are therefore necessary. Item separation indicates the extent to which the test items can be separated in terms of difficulty; an item separation index less than 3.0 with a person reliability coefficient less than 0.9 suggests that the sample of test takers is too small to distinguish items of high and low difficulty. Whereas person reliability is influenced by variance of ability in the sample, test length, number of categories per item, and sample–item targeting, item difficulty is influenced by variance of item difficulty and sample size. Therefore, to improve person reliability, it is necessary to have a wider range of ability among test takers, numerous test items and categories, and better targeting. To improve item reliability, it is also necessary to have a wide range of difficulty among test items and a large sample size. Since goals of the study include assessing a wide variety of proficiency levels of items and not dividing test takers by performance, low person reliability and person separation index were excluded as criteria.

Test targeting gauges the extent to which the difficulty of test items is appropriate for a test taker's estimated level of ability. Winsteps provides a distribution of item difficulty and person ability estimates on the same continuum of measurement unit in order to allow their comparison. It also provides the most probable response key map, which positions the difficulty metric along the x-axis. Items are listed along the right side of the figure, with most difficult items at the top. A horizontal gap between the items indicates that no items covered that space. The map thus provides insights into which levels of difficulty the test lacks.

Winsteps offers infit and outfit mean square fit statistics and principal components analysis of residuals [20]. Item fit statistics provide information regarding the extent to which the observed test taker's response corresponds to the expected response based on the Rasch model. Winsteps also provides two types of fit statistics to assess the residual difference between actual and expected responses. On the one hand, the infit mean square statistic (MNSQ) is affected by unexpected responses such as a high performer mistakes on an item of low difficulty; on the other, outfit MNSQ is affected by any unexpected pattern of responses proximate to a test taker's ability estimates. [20] has provided a table of values and their meanings, in which values between 0.5 and 1.5 are assumed to be productive measurements. The present research adopted [21]'s guidelines for item fit; items with MNSQ values greater than 1.4 with ZSTDs in excess of 2.0 were categorized as misfitting, whereas items with MNSQ values less than 0.6 with ZSTD values less than −2.0 were categorized as overfitting. Items that fit those guidelines were analyzed for their content and assessed whether the value improved when the item was omitted.

[20] has suggested that the aim of principal components analysis of Rasch residuals is "to extract the common factor that explains the most residual variance under the hypothesis that there is such a factor." When the persons and items fit the model expectation, the study examined whether variance explained by the first contrast was less than 10% [22].

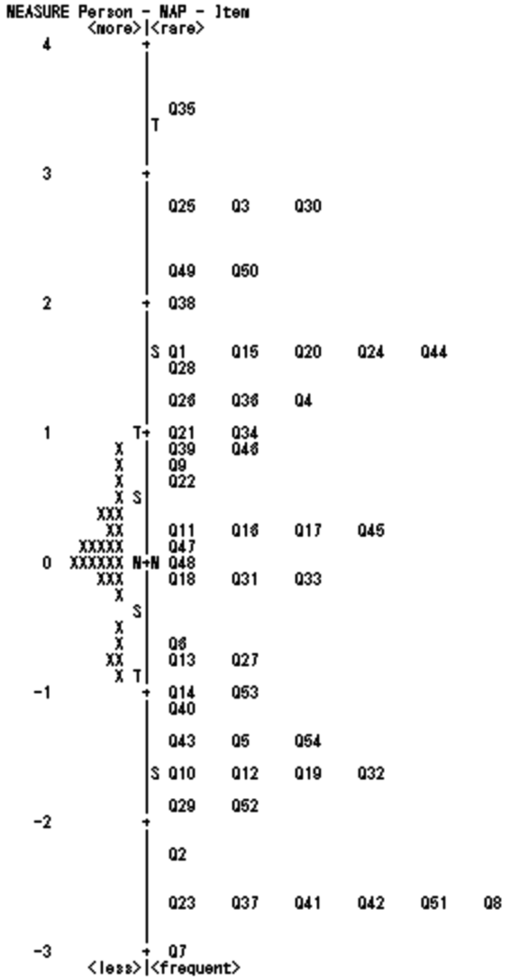
Results and Discussion

Results of the Rasch analysis of 54 test items of a final test in 2015 shown below reveal the statistical characteristics of the test and the problems needed to be resolved in its revision.

Original Analysis

The separation measure concerning the original analysis of the final test in 2015 is relatively low (0.77–0.85), which indicates that the number of items used is rather small to distinguish persons. Its person reliability is low (0.41), which is due to a relatively narrow range of person measure compared to that of item measure. The reliability of item is moderately high (0.89) which indicates that, if the items were given to other comparable groups of test takers, there is a probability that the test would reproduce a similar order of item hierarchy. The item separation measure of 2.83–2.87 indicates that the items can be separated into more than two strata of difficulty [23].

Figure 1 shows that the persons mean (0.04) is located slightly above the items mean that is set to 0.00 by default, indicating that, on average, items are approximately



```
MEASURE   Person - MAP - Item
           <more>|<rare>
    4            +
                 |
                 |
                 |
                 |  Q35
                 |T
                 |
                 |
    3            +
                 |
                 |  Q25    Q3     Q30
                 |
                 |
                 |
                 |  Q49    Q50
                 |
    2            +  Q38
                 |
                 |
                 |S Q1     Q15    Q20    Q24    Q44
                 |  Q28
                 |
                 |  Q26    Q36    Q4
                 |
    1           T+  Q21    Q34
              X  |  Q39    Q46
              X  |  Q9
              X  |  Q22
              X S|
            XXX  |
             XX  |  Q11    Q16    Q17    Q45
          XXXXX  |  Q47
    0    XXXXXX M+M Q48
            XXX  |  Q18    Q31    Q33
              X  |
                S|
              X  |
              X  |  Q6
             XX  |  Q13    Q27
              X T|
   -1            +  Q14    Q53
                 |  Q40
                 |
                 |  Q43    Q5     Q54
                 |
                 |S Q10    Q12    Q19    Q32
                 |
                 |  Q29    Q52
   -2            +
                 |
                 |  Q2
                 |
                 |
                 |  Q23    Q37    Q41    Q42    Q51    Q8
                 |
                 |
   -3            +  Q7
           <less>|<frequent>
```

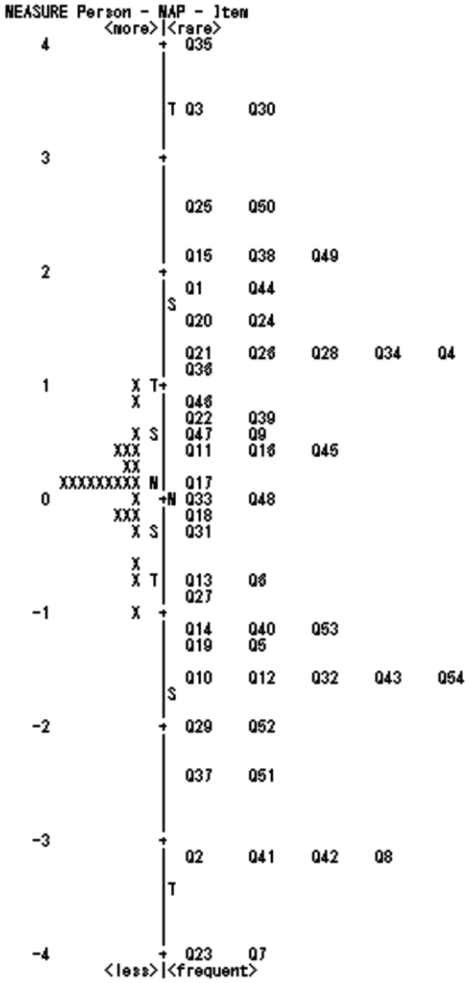
Figure 1. *Item and person map for the final test in 2015 (original analysis).*

at the right level for persons. Some item difficulty estimates, however, fall far below the person ability estimates for the final test in 2015.

Of the 29 persons measured, all MNSQ ranges fell between 0.6–1.4, except for persons 9 (Infit MNSQ:1.52, ZSTD:2.51), 10 (Infit MNSQ:1.41, ZSTD:2.10), 19 (Infit MNSQ:0.55, ZSTD:−3.04), and 23 (Outfit MNSQ:2.33, ZSTD:2.86)'s MNSQ value with a ZSTD value, possibly indicating mismatched persons. As [21] has pointed out, an MNSQ value exceeding 1.4 with a ZSTD value more than 2.0 or an MNSQ value less than 0.6 with a ZSTD value less than −2.0, indicates a possibly mismatched person, suggesting that persons 9, 10, 19, and 23 should be deleted from the list. Of the 54 items measured, all MNSQ ranges fell between 0.6–1.4.

1st revision

As shown above, there were 54 items and 25 persons in the 1st revision. WINSTEPS version 3.81.0 provides the statistical values of the test items of 1st revision. The separation measure concerning the 1st revision of the final test in 2015 is low (0.61–0.68), which indicates that the number of items used is rather small to distinguish persons. Its person reliability is very low (0.3), which is due to the relatively small number of test items. The reliability of item is moderately high (0.86–0.87) which indicates that, if the items were given to other comparable groups of test takers, there is a probability that the test would reproduce a similar order of item hierarchy. The item separation measure of 2.53–2.56 indicates that the items can be separated into



```
MEASURE Person - MAP - Item
          <more>|<rare>
    4           +  Q35
                |
                |
                |
                |T Q3     Q30
                |
                |
    3           +
                |
                |
                |  Q25    Q50
                |
                |
                |  Q15    Q38    Q49
    2           +
                |  Q1     Q44
                |S
                |  Q20    Q24
                |
                |  Q21    Q26    Q28    Q34    Q4
                |  Q36
    1         X T+
             X  |  Q46
                |  Q22    Q39
             X S|  Q47    Q9
           XXX  |  Q11    Q16    Q45
            XX  |
     XXXXXXXXX N|  Q17
    0        X  +N Q33    Q48
           XXX  |  Q18
             X S|  Q31
                |
             X  |
             X T|  Q13    Q6
                |  Q27
   -1        X  +
                |  Q14    Q40    Q53
                |  Q19    Q5
                |
                |  Q10    Q12    Q32    Q43    Q54
                |S
                |
   -2           +  Q29    Q52
                |
                |
                |  Q37    Q51
                |
                |
                |
   -3           +
                |  Q2     Q41    Q42    Q8
                |
                |T
                |
                |
                |
   -4           +  Q23    Q7
          <less>|<frequent>
```

Figure 2. *Item and person map for the final test in 2015 (1^{st} revision).*

Table 2 *Standardized Residual Variance (in Eigen-value Units) for the Final Test in Fall*

	Observed			Expected
Total raw variance in observations	83.3	100%		100%
Raw variance explained by measures	32.3	38.8%		38.8%
Raw variance explained by persons	3.2	3.8%		3.8%
Raw variance explained by items	29.1	35.0%		34.9%
Raw unexplained variance (total)	51.0	61.2%	100%	61.2%
Unexplained variance in 1st contrast	5.6	6.8%	11.0%	
Unexplained variance in 2nd contrast	4.7	5.6%	9.2%	
Unexplained variance in 3rd contrast	4.4	5.3%	8.7%	
Unexplained variance in 4th contrast	4.2	5.0%	8.2%	
Unexplained variance in 5th contrast	4.0	4.8%	7.8%	

Note. This figure is from WINSTEPS® version 3.81.0 output Table 23.0.

more than two strata of difficulty [23].

Figure 2 shows that the persons mean (0.12) is located just above the items mean that is set to 0.00 by default, indicating that, on average, items are at an appropriate level of difficulty for persons. Some item difficulty estimates, however, fall far below the person ability estimates for the final test in 2015. Of the 25 persons measured, all MNSQ ranges fell between 0.6-1.4. Of the 54 items measured, all MNSQ ranges fell between 0.6-1.4.

Table 2 indicates that observed raw variance explained by measures (38.8%) fits the expected raw variance explained by measure (38.8%), indicating that explainable variance fits the Rasch model. Rasch, however, explained only 35% of the 54 items, leaving more than half of variance (61.2%) unaccounted for by the model. This is due to the fact that the ability range of the test takers is relatively narrow. A wider proficiency level of the test takers would result in a greater explained variance. The strongest secondary dimension is named the first contrast, while the following dimensions are named the second, third, fourth and fifth respectively. The largest secondary dimension (first contrast) in the final test in 2015 data had strength of 5.6 units (6.8%) while the variance explained by measures was larger at 32.3 units (38.8%), indicating that the secondary contrast does not create multidimensionality.

References

[1] Ehri, L.C., & Soffer, A.G. (1999). Graphophonemic awareness: Development in elementary students. *Scientific Studies of Reading, 3*(1), 1-30.

[2] Connelly, V. (2002). Graphophonemic awareness in adults after instruction in phonic generalisations. *Learning and Instruction, 12*(6), 627-649.

[3] Compton, D.L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4*(3), 219-259.

[4] Nagy, W.E., Kuo-Kealoha, A., Xinchun, W., Wenling, L., Anderson, R.C., & Xi, C. (2002). The role of morphological awareness in learning to read Chinese. In *Chinese children's reading acquisition* (pp. 59-86). Springer US.

[5] Thompson, G.B., Connelly, V., Fletcher-Flinn, C.M., & Hodson, S.J. (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Memory & Cognition, 37*(2), 223-234.

[6] Szabo, S. (2010). Older children need phonemic awareness instruction, too. *TESOL Journal, 1*(1), 130-141.

[7] Scarborough, H.S., Ehri, L.C., Olson, R.K., & Fowler, A.E. (1998). The fate of phonemic awareness beyond the elementary school years. *Scientific Studies of Reading, 2*(2), 115-142.

[8] Ichiyama, Y. (2018). Orthographic and phonological features of English textbooks in junior high school in Japan. *International Journal of Science and Research, 7*(1), 372-374.

[9] Suski, P.M. (1931). The phonetics of Japanese language. *Routledge library edition ‖ 59.*

[10] Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2011). The Course of Study for senior high schools guidelines explanation. Retrieved on September 2017, from http://www.mext.go.jp/a_menu/shotou/newcs/youryou/eiyaku/__icsFiles/afieldfile/2011/04/11/1298353_9.pdf

[11] Kay, G. (1995). English loanwords in Japanese. *World Englishes, 14*(1), 67-76.

[12] Ohata, K. (2004). Phonological differences between Japanese and English: Several potentially problematic areas of pronunciation for Japanese ESL/EFL learners. *The Asian EFL Journal, 6*(4), 1-19.

[13] McBride, J.R., Ysseldyke, J., Milone, M., & Stickney, E. (2010). Technical adequacy and cost benefit of four measures of early literacy. *Canadian Journal of School Psychology, 25*(2), 189-204.

[14] Missall, K., Reschly, A., Betts, J., & McConnell, S. (2007). Examination of the predictive validity of preschool early literacy skills. *School Psychology Review, 36*(3), 433.

[15] Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing, 27*(1), 101-118.

[16] Haladyna, T.M. (2012). *Developing and validating multiple-choice test items*. Routledge.

[17] Baghaei, P., & Amrahi, N. (2011). Validation of a Multiple Choice English Vocabulary Test with the Rasch Model. *Journal of Language Teaching & Research, 2*(5).

[18] Linacre, J.M. (2006). WINSTEPS Rasch measurement computer program (version 3.81.0). Chicago: Winsteps.com

[19] Linacre, J.M. (2012). Winsteps® Rasch Tutorial 4. Available at: http://www.winsteps.com/a/winsteps-tutorial-4.pdf

[20] Linacre, J.M. (2013). A User's Guide to WINSTEPS R. Ministep: Rasch-model computer programs [program manual 3.80.0]. Chicago: IL.

[21] Jarl, G.M., Heinemann A.W., & Hermansson L.M.N. (2012). Validity evidence for a modified version of the orthotics and prosthetics users' survey. *Disability and Rehabilitation: Assistive Technology*, 7, 469-478.

[22] Fischer, W.P.J. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transaction, 21*(1). 1095.

[23] Karim, A.A., Shah, P.M., Din, R., Ahmad, M., & Lubis, M.A. (2014). Developing information skills test for Malaysian youth students using Rasch analysis. *International Education Studies, 7*(13), 112-122.

Acknowledgement:

This work was supported by Grant-in-Aid for Scientific Research (C) Research No. 15K02800. I am grateful for their support.