四天王寺大学紀要　第69号（2021年3月）

Validating the *Lunic Language Marathon* as a Feasible Measurement for L2 Aptitude

第二言語適性測定のための有効な指標としてルニック・ランゲージ・マラソンを検証

Robert J. Kerrigan
ケリガン ロバート

Abstract:

Foreign language (L2) aptitude has been recognized as a cognitive construct that is monumental in a person's ability to learn a foreign language. Researching aptitude once held prominence in the field of second language acquisition, but it has received less attention from researchers of late. However, some researchers have acknowledged its significance in the role L2 acquisition, and the construct of L2 aptitude has evolved to better highlight its complex nature. Furthermore, there are a plethora of instruments designed to accurately measure a leaner's L2 aptitude, such as the MLAT and the PLAB. However, such instruments can be taxing on a learner's cognitive resources. The purpose of this study was to assess and validate one type of L2 aptitude test, known as the *Lunic Language Marathon* (LLM), as a more accessible viable option compared to more well-known aptitude tests. In this study, 121 university students participated in the validation study by taking the first two subsections of the LLM: the *Lunic Numbers* and *Lunic Writing*. The results were analyzed using Rasch analysis. The findings revealed that as a general measurement of L2 aptitude, the LLM is sufficient. Furthermore, a closer inspection of the *Lunic Numbers* subsection indicated that it was a reliable measurement of auditory memory. However, the *Lunic Writing* subsection suffered from a ceiling effect, affecting the reliability of this section as a measurement of phonetic coding ability.

Keywords:
L2 aptitude, Lunic Language Marathon, auditory memory, phonetic coding ability, Rasch analysis, instrument validation

Foreign language (L2) aptitude has been recognized as an individual difference variable that is instrumental in one's ability to acquire an additional language beyond the critical period. Originally considered a single, unidimensional cognitive construct—you either have it, or you do not—modern conceptualizations of the construct construe it as a multidimensional construct. That is, L2 aptitude can refer to different aspects of learning a language, such as having a knack for learning vocabulary or discriminating L2 phonemes. Various researchers have developed instruments to measure learners' L2 aptitude in order to ascertain their ability to learn a foreign language with relatively high levels of success. The purpose of this study is to validate a lesser-known aptitude test, the *Lunic Language Marathon* (LLM), as a viable instrument for measuring Japanese students' L2 aptitude. This study begins with an exploration of the L2 aptitude construct, followed by a review of aptitude treatment studies. The remaining sections of the study delve into the validation study of the LLM, including a discussion of the results of the validation process.

Aptitude as a Construct

L2 aptitude is considered a cognitive trait, which is “characterized as strengths individual learners have—relative to their population—in the cognitive abilities [in which] information processing draws on during L2 learning and performance in various contexts and at different stages” (Robinson, 2005, p. 46). Language aptitude not only depicts the degree of strength or relative ease of learning a language, it is also concerned with the rate of progress—a person with strong language aptitude should learn an L2 quicker than someone with average or low L2 aptitude regardless of instructional setting (Dörnyei & Ryan, 2015; Kiss & Nikolov, 2005). It has also been described as a talent for learning languages (Skehan, 1998) and as an inherent foreign language learning ability (Ortega, 2009). Skehan (1998) asserted that language aptitude is a construct, as it is distinct from general aptitude or general intelligence, it is stable over time, and it is inherent and untrainable. Aptitude and, more specifically, language aptitude have been shown to be related to intelligence but are distinct constructs. Sasaki (1993) was able to provide evidence through her structural equation model that intelligence and aptitude were distinct second-order latent variables as part of the first-order latent variable of cognition.

Other aptitude researchers have noted that cognition is just one aspect of the triad of aptitude (the other two being affective and conative factors: Robinson, 2007). Earlier work on aptitude research assumed that it was monolithic in nature; that is, researchers assessed L2 aptitude as a general trait (Skehan, 2002). However, it has been acknowledged that language aptitude needs to be regarded as componential in nature. In other words, aptitude is not a unidimensional cognitive construct in which someone is strong in it or not, but rather, there are various aspects of language aptitude, where individuals can be strong in one aspect but weak in others. Robinson (2001, 2002a, 2002b, 2005, 2007) labelled this as aptitude complexes. This componential conceptualization of language aptitude is considered hierarchical in nature.

Studies on Aptitude and Interaction Effects

Aptitude treatment interaction (ATI) studies base their treatment groups on the aptitude profiles of the participants (Erlam, 2005; Skehan, 1998; Snow, 1991). There have been very few researchers that have looked at how programs can be tailored to suit the needs of the learners based on their aptitude profile. Wesche (1981) was the one of the first and few researchers to investigate the interaction effects of aptitude in relation to the approach of instruction. In the study, the participants were placed in either an audio-visual approach, analytical approach, or functional approach groups based on their scores from the Modern Language Aptitude Test (MLAT: Carroll & Sapon, 1959) and the sound discrimination and sound-symbol association subtests of the Pimsleur Language Aptitude Battery (PLAB: Pimsleur, 1966). Furthermore, a group of participants were mismatched into either the audiovisual or analytic treatments, and it was found that those who were matched into their appropriate treatment as determined by their aptitude test scores showed greater achievement on three out of four achievement tests, more interest in learning languages, more enthusiasm to study the language autonomously, a more positive attitude towards their instructional approach, and less L2 anxiety. Clearly, this study provided evidence that treatments that are tailored to suit the learning styles of students can have positive benefits on their L2 learning.

Although ATI studies that have been designed based on the aptitude profiles of the participants have been rare since the Wesche (1981) study, there have been researchers who have conducted studies focusing on how different components of language aptitude can affect various aspects of L2 acquisition. Robinson's (1997) study on implicit and explicit instruction in relation to the individual difference variables of aptitude and participants' abilities to be aware of linguistic cues in the input created a resurgence in L2 aptitude research. The results of the study provided evidence for the componential nature of aptitude and that those who excel in various forms of instructional conditions do so based on their aptitude profile. Although not an ATI study in the sense that the treatment was specifically designed around aptitude, the results from the Erlam (2005) study gave further credence to aptitude being a multidimensional construct. Furthermore, Erlam stated those with strong analytical abilities benefitted from implicit instruction, whereas aptitude effects were negated for participants who received deductive instruction. In other words, explicit instruction was beneficial for those who tended to have weak language aptitude.

Sheen (2007) contradicted such findings, declaring that aptitude is instrumental in deductive approaches. Sheen examined how aptitude can influence a learner's ability to take advantage of corrective feedback. Sheen gave corrective feedback on written article use and used two experimental treatments, a direct-only correction treatment (n=31) and a direct metalinguistic correction treatment (n=32), and a control group (n=28). To test participants' language aptitude, Sheen used a language ability test, similar to the language analysis test of the PLAB. The results from the ANCOVA suggested that in order to benefit from correctional feedback, particularly metalinguistic feedback, the participants needed to have strong analytic abilities. Although this study highlighted the componential role aptitude can play in corrective feedback, these results cannot be fully compared to Robinson (1997) or Erlam (2005), as Sheen did not include an implicit or incidental group to determine if language analytic aptitude has a role in noticing incidental or implicit corrective feedback, and measures for cognitive constructs were collected exogenously, creating an issue of recursiveness. Furthermore, a regression analysis could have provided insight into how much variation language analytic aptitude contributes to the learning of articles via corrective feedback.

In terms of ATI studies, age and the critical period have been instrumental in assessing the importance of aptitude on L2 learning. Ross, Yoshinaga, and Sasaki (2002) investigated how the critical age period could explain the degree of intuitiveness of correct instances of wh-movement with a group of Japanese learners when other individual difference variables, such as aptitude, were taken into consideration. The participants represented three different time periods for their age of onset and learning context:

naturalistic exposure from childhood (*n*=34), immersion from their teens (*n*=38), and exclusive classroom-based foreign language teaching after the critical period (*n*=57). These experimental groups were also compared with an English-native speaker group. The participants were given a wh-movement violation survey, which resembled a grammaticality judgement test (GJT). The participants were presented with 24 sentences and had to indicate whether the sentences were acceptable or not on a seven-point scale. They also took the words in sentences subtest of the MLAT and the artificial language analysis subtest of the Language Aptitude Battery for the Japanese (LABJ: Sasaki, 1996). An ANCOVA analysis with the results of the sentences with correct use of wh-movement as the dependent variable (DV) and the aptitude measurements as the covariates revealed that the child SLA group performed the closest to the native-speaker group than the other experimental groups. Furthermore, Ross et al. examined interaction effects of age of onset with aptitude and found that those with higher aptitude scores could compensate for their lack of L2 exposure and identify wh-movement violations similarly to those who were immersed in the L2 from a young age. They concluded that aptitude is at its most important in post-critical stages of learning, whereas it plays no significant role for young learners.

In a study similar to Ross et al. (2002), Harley and Hart (1997) were also concerned with how age of onset and aptitude profiles can affect L2 acquisition. The first group (*n*=36) started partial French immersion from Grade 1, and the second group (*n*=29) started partial French immersion from Grade 7. They tested the participants' aptitude via the word pairs subtest from the MLAT, the language analysis subtest from the PLAB, and a memory-for-text measure. To assess proficiency, they used vocabulary recognition, listening comprehension, cloze, written production, and oral tests. Stepwise regressions revealed that the memory-for-text test was significant for the vocabulary recognition, listening comprehension and cloze tests for the early immersion group. For the late immersion group, the language analysis subtest was a significant predictor for the vocabulary recognition, cloze tests and the two measurements for the written production tests. The results of this study indicated, as in the Ross et al. study, that learners rely on different components of aptitude when different ages of onset are taken into account. Learners past the critical period tend to use their analytical skills to attend to linguistic processing, whereas younger learners depend on memory systems of what they have previously encountered to process language. This study provided further evidence that aptitude is componential in nature and taking a componential view of aptitude can facilitate in how to approach instructional contexts.

Piloting the Lunic Language Marathon

The LLM was developed by Sick and Irie (2001) as an aptitude test designed for diagnosis and research purposes (Sick, 2007). This test has advantages for testing the language aptitude of Japanese learners over other well-known aptitude tests, such as the MLAT, for three reasons. First, it is more applicable to Japanese learners of English because the instructions are written and recorded in Japanese. Second, an artificial language is used, removing the confound of the respondents not being dependent on their knowledge of the English language (Kiss & Nikolov, 2005). Finally, the test is formatted as an activity to distract test-takers from thinking that they are taking an aptitude test. Using this type of format might be more appealing to participants than a traditional test, such as the MLAT, and there is less risk of test fatigue.

The LLM is comprised of five subsections: *Lunic Numbers*, *Lunic Writing*, *Lunic Vocabulary*, *Lunic Grammar*, and a survey about the test-takers' impressions of the various tasks in the LLM and their ability to think introspectively about whether they approached the subtests analytically or by using memory strategies (Sick & Irie, 2001). Because I aim to investigate whether phonetic coding improves over a long treatment period of listening practice or not for future research, I only used the first two subsections of the LLM. In the *Lunic Numbers* subsection, test-takers are given a short aural lesson of the *Lunic* number system, and they then do a dictation task. They will listen to 15 items of three-digit *Lunic* numbers and write down the numbers they hear. They are awarded a correct score for any digit they write down correctly. For example, if a participant hears the number 413 in *Lunic* and writes down 403, that participant will get two out of three points for that item. This section is based on the number learning section of the MLAT and is designed to test auditory memory and auditory learning ability. In the *Lunic Writing* subsection, test-takers learn the orthographic system of *Lunic* and the phonetic forms of the *Lunic* characters. Afterwards, the test-takers listen to a series of *Lunic* letters and choose their correct orthographic form. There are four sections for this subtest consisting of five items each. For each section, they listen to each *Lunic* word in each item. Afterwards, they hear one word for each item and have to indicate which word on the answer sheet corresponds to the word they heard. They receive one point for each correct answer. This section is based on the phonetic script subtest of the MLAT and is designed to test phonetic coding ability through sound-symbol associations.

Participants

For the pilot study, 121 university students participated in the study. The participants came from a small private university in western Japan. Sampling was done via nonprobability convenience sampling (6 intact classes). Ages ranged from 18 to 21 (M=19.74, SD=.73). The number of years of formal English study varied from seven to ten years. A number of students had studied abroad prior to this study from between one month to five months (study abroad data unavailable).

Procedure

I administered the *Lunic* numbers and *Lunic* writing subtests of the LLM to the participants during the latter half of class time. Due to unavoidable circumstances, I had to administer the test for two classes after they had completed their mid-term test, which could have led to test fatigue. I explained to each class that we were going to do a language learning challenge and that I was not testing their language learning ability but rather the efficacy of the test. I explained the procedures in English but that they would then listen to the instructions in Japanese. Finally, I instructed them to do the task by themselves and that I would collect their answer sheets after the activity. The tests for each class took approximately 25 minutes to complete.

Analysis

The participants' dichotomous responses were input into a control file and then analyzed using Winsteps 4.0.0 (Linacre, 2009). I analyzed the data using the dichotomous Rasch model. This model predicts the possibility of a person being able to answer an item correctly given their ability level and the item's difficulty. To get person estimates of ability, the raw score percentages are converted into odds of success. This is done by calculating the ratio of a participant's percentage of correct responses over the percentage of incorrect responses and the natural log for each person. Similarly, to calculate item difficulty, an item is divided by the number of people who answered it correctly with the number of people who incorrectly answered it and taking the natural log of that value. Both items and persons can now be expressed on a scale of log odds ratio (or logits) that is standardized and is now interval-level data (Bond & Fox, 2015). With this conversion, the difficulty of an item in relation to a respondent can be determined and vice versa. For instance, a person with a logit of -1.0 would have a very low probability of correctly answering an item with a logit of 2.0.

Furthermore, with this model, researchers can identify either items or persons that are not performing as expected and eliminate them from the analysis or, in the case of items, revise them, as indicated by their fit. To identify potential misfitting items or participants, I set the range from 0.7 to 1.30 for infit and outfit statistics based on the recommendations of Bond and Fox (2015). If an item does not meet those parameters for fit, it indicates overfit, meaning that responses follow a Guttman pattern; that is, there is more consistency than predicted according to the Rasch model. If it exceeds the parameters, there is underfit in which case, there is more variation in responses to the item than what the Rasch model predicted. Likewise, if a person has a fit statistic of under 0.7, it indicates that the person is responding to items too systematically and are performing better than the Rasch model expected (Bond & Fox, 2015). If there are such instances of misfit for persons or items, they can be checked to see if there is a fault with the item or the person was not taking the test seriously.

Results and Analyses

I analyzed the data in three separate analyses. The first analysis was with the composite aptitude data of both LLM subtests. For the second and third analyses, I analyzed the data from the *Lunic* numbers and *Lunic* writing subtests respectively. The responses were analyzed for their reliability and goodness-of-fit for person and item responses.

The composite LLM results. The results of the analysis for the combined scores of the *Lunic* numbers and *Lunic* writing subtests revealed strong person reliability (.86) with two participants removed from the analysis due to not completing the instruments. Person separation was 2.63, which indicated that the analysis could reliably divide the participants into at least two statistically distinct groups. Item reliability was very strong (.90). The results for person analysis can be seen in Table 1. Item dimensionality analysis showed one item with borderline infit (1.31). However, removal of that item made no changes to reliability, so it was retained in the analysis. Item separation was 3.04, suggesting that the sample provided a good range of difficulty of items. The results can be seen in Table 2. The Wright map (see Figure 1) shows the spread of items and persons. The map shows that the range of person abilities was well-matched to the range of item difficulties, indicating that the LLM is well targeted to this sample of learners. Item dimensionality revealed that the items explained 17.7 percent of the variance with 13.98 eigenvalues. The first contrast had an Eigenvalue of 4.53, and the second one was 3.61. This suggested that the instrument might not be unidimensional. However, that is understandable, as it was the composite analysis of the LLM. In order to determine if each of the subtests were reliable and unidimensional, I analyzed them separately.

Table 1
Summary Table for Person Measures for the Composite Results of the LLM

	Total Score	Measure	Real *SE*	Infit *MNSQ*	Infit *ZSTD*	Outfit *MNSQ*	Outfit *ZSTD*
M	30.90	-.12	.28	1.00	.00	1.01	.00
P. *SD*	10.20	.76	.03	.11	1.20	.18	1.30
S. *SD*	10.30	.77	.03	.11	1.20	.18	1.30
Max	54.00	1.73	.47	1.27	3.40	1.93	3.30
Min	5.00	-2.66	.26	.75	-2.70	.63	-2.60
REAL *RMSE*	.28	TRUE *SD*	.71	SEP	2.51	PER REL	.86
MODEL *RMSE*	.28	TRUE *SD*	.71	SEP	2.57	PER REL	.87
SE OF PERSON MEAN			.08				

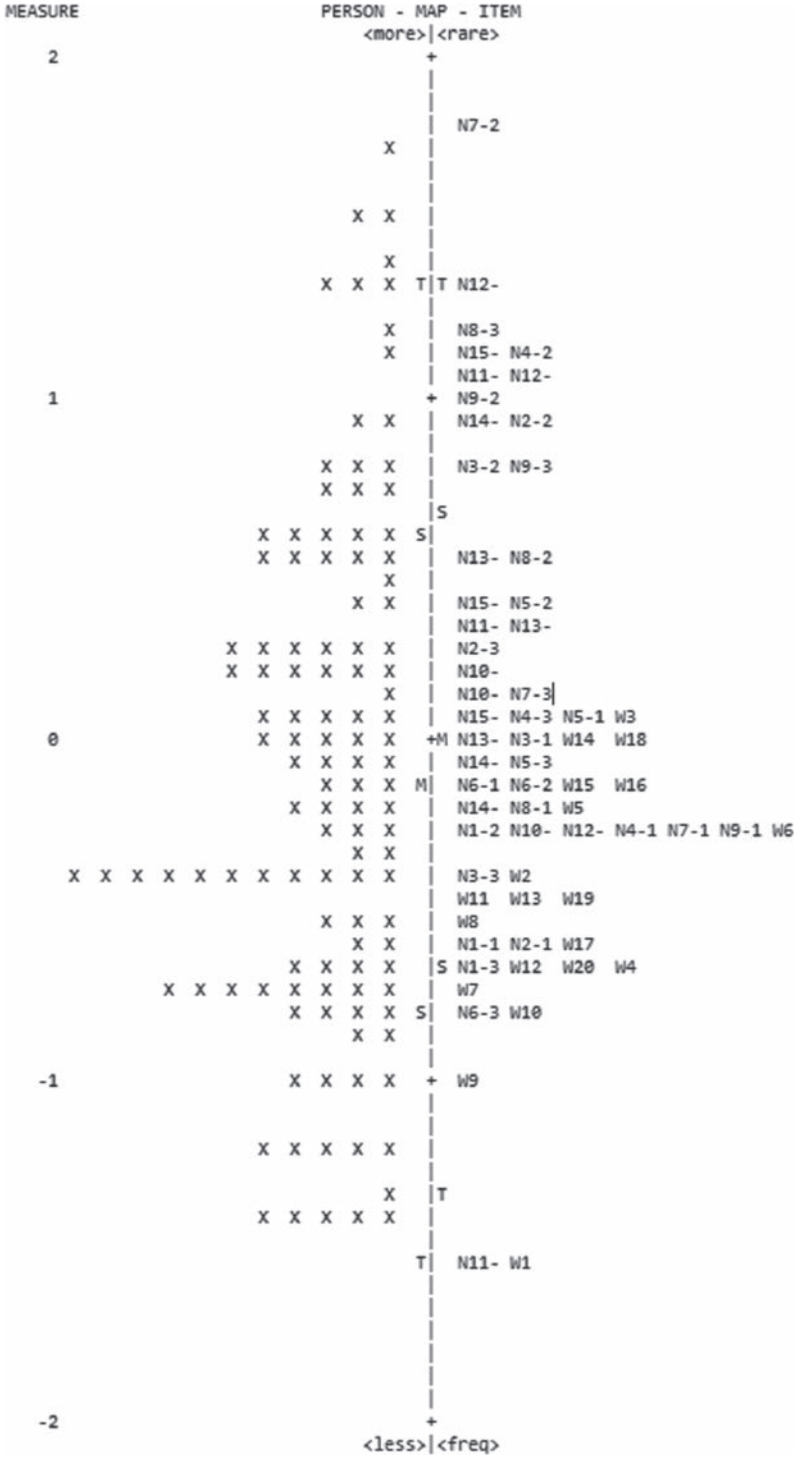
Note. SD = Standard deviation; P. SD = Population standard deviation; S. SD = Sample standard deviation; SE = Standard error; Max. = maximum value; Min = minimum value; MNSQ = mean-squared; ZSTD = Standardized z-scores; RSME = square-root of the average error variance; SEP = Separation; PER REL. = Person reliability.

Table 2
Summary Table for Item Measures for the Composite Results of the LLM

	Total Score	Measure	Real *SE*	Infit *MNSQ*	Infit *ZSTD*	Outfit *MNSQ*	Outfit *ZSTD*
M	56.60	.00	.21	1.00	-1.00	1.01	-1.00
P. *SD*	16.2	.67	.02	.10	1.30	.15	1.40
S. *SD*	16.3	.67	.02	.10	1.30	.15	1.40
Max	93.00	1.79	.29	1.31	3.80	1.54	4.50
Min	48.00	-1.35	.23	.80	-2.30	.73	-2.20
REAL RMSE	.21	*TRUE SD*	.63	SEP	3.04	ITEM REL	.90
MODEL RMSE	.20	*TRUE SD*	.63	SEP	3.11	ITEM REL	.91
SE OF PERSON MEAN			.08				

Note. SD = Standard deviation; P. SD = Population standard deviation; S. SD = Sample standard deviation; SE = Standard error; Max. = maximum value; Min = minimum value; MNSQ = mean-squared; ZSTD = Standardized z-scores; RSME = square-root of the average error variance; SEP = Separation; ITEM REL = Item reliability.

Figure 1
The Wright Map for Items and Persons for the Composite Results of the LLM



```
MEASURE                       PERSON - MAP - ITEM
                                  <more>|<rare>
    2                                   +
                                        |
                                        |
                                        |  N7-2
                                    X   |
                                        |
                                        |
                                 X  X   |
                                        |
                                    X   |
                              X  X  X  T|T N12-
                                        |
                                    X   |  N8-3
                                    X   |  N15- N4-2
                                        |  N11- N12-
    1                                   +  N9-2
                                 X  X   |  N14- N2-2
                                        |
                              X  X  X   |  N3-2 N9-3
                              X  X  X   |
                                        |S
                        X  X  X  X  X  S|
                        X  X  X  X  X   |  N13- N8-2
                                    X   |
                                 X  X   |  N15- N5-2
                                        |  N11- N13-
                     X  X  X  X  X  X   |  N2-3
                     X  X  X  X  X  X   |  N10-
                                    X   |  N10- N7-3
                        X  X  X  X  X   |  N15- N4-3 N5-1 W3
    0                   X  X  X  X  X   +M N13- N3-1 W14  W18
                           X  X  X  X   |  N14- N5-3
                              X  X  X  M|  N6-1 N6-2 W15  W16
                           X  X  X  X   |  N14- N8-1 W5
                              X  X  X   |  N1-2 N10- N12- N4-1 N7-1 N9-1 W6
                                 X  X   |
      X  X  X  X  X  X  X  X  X  X  X   |  N3-3 W2
                                        |  W11  W13  W19
                              X  X  X   |  W8
                                 X  X   |  N1-1 N2-1 W17
                           X  X  X  X   |S N1-3 W12  W20  W4
               X  X  X  X  X  X  X  X   |  W7
                           X  X  X  X  S|  N6-3 W10
                                 X  X   |
                                        |
   -1                      X  X  X  X   +  W9
                                        |
                                        |
                        X  X  X  X  X   |
                                        |
                                    X   |T
                        X  X  X  X  X   |
                                        |
                                       T|  N11- W1
                                        |
                                        |
                                        |
                                        |
                                        |
                                        |
   -2                                   +
                                  <less>|<freq>
```

Analysis of the *Lunic* numbers subtest. The results for the person estimates showed person reliability to be .80 and person separation at 1.99 (see Table 3 for the summary of the person estimate results). The results for items showed item reliability to be .90 and item separation at 2.98, which indicated that there was an adequate range of person ability to reliably estimate the item difficulties (see Table 4 for the summary of the item estimate results). Analysis for misfit did not showed any seriously underfitting or overfitting items, so all items were retained. The Wright map for the *Lunic* numbers subtest provided evidence that the items and persons were well-matched.

Item dimensionality analysis showed that the *Lunic* numbers subtest had an eigenvalue of 9.78, which explained 17.8 percent of the variance. The first contrast had an eigenvalue of 3.98 (6.9% of the variance), and the second contrast had an eigenvalue of 2.67 (4.9%). This might suggest that there is an underlying variable that is not being accounted for in the measurement.

Table 3

Summary Table for Person Measures for the Results of the Lunic Numbers Subtest of the LLM

	Total Score	Measure	Real *SE*	Infit *MNSQ*	Infit *ZSTD*	Outfit *MNSQ*	Outfit *ZSTD*
M	19.40	-.34	.34	1.00	-1.00	1.02	.00
P. SD	7.00	.76	.03	.11	1.00	.18	1.10
S. SD	7.10	.77	.03	.11	1.00	.18	1.10
Max	34.00	1.23	.49	1.27	2.30	1.72	2.80
Min	5.00	-2.24	.31	.78	-2.70	.73	-2.60
REAL RMSE	.34	*TRUE SD*	68	SEP	1.99	PER REL	.80
MODEL RMSE	.34	*TRUE SD*	.68	SEP	2.04	PER REL	.81
SE OF PERSON MEAN			.07				

Note. SD = Standard deviation; P. SD = Population standard deviation; S. SD = Sample standard deviation; SE = Standard error; Max. = maximum value; Min = minimum value; MNSQ = mean-squared; ZSTD = Standardized z-scores; RSME = square-root of the average error variance; SEP = Separation; PER REL. = Person reliability.

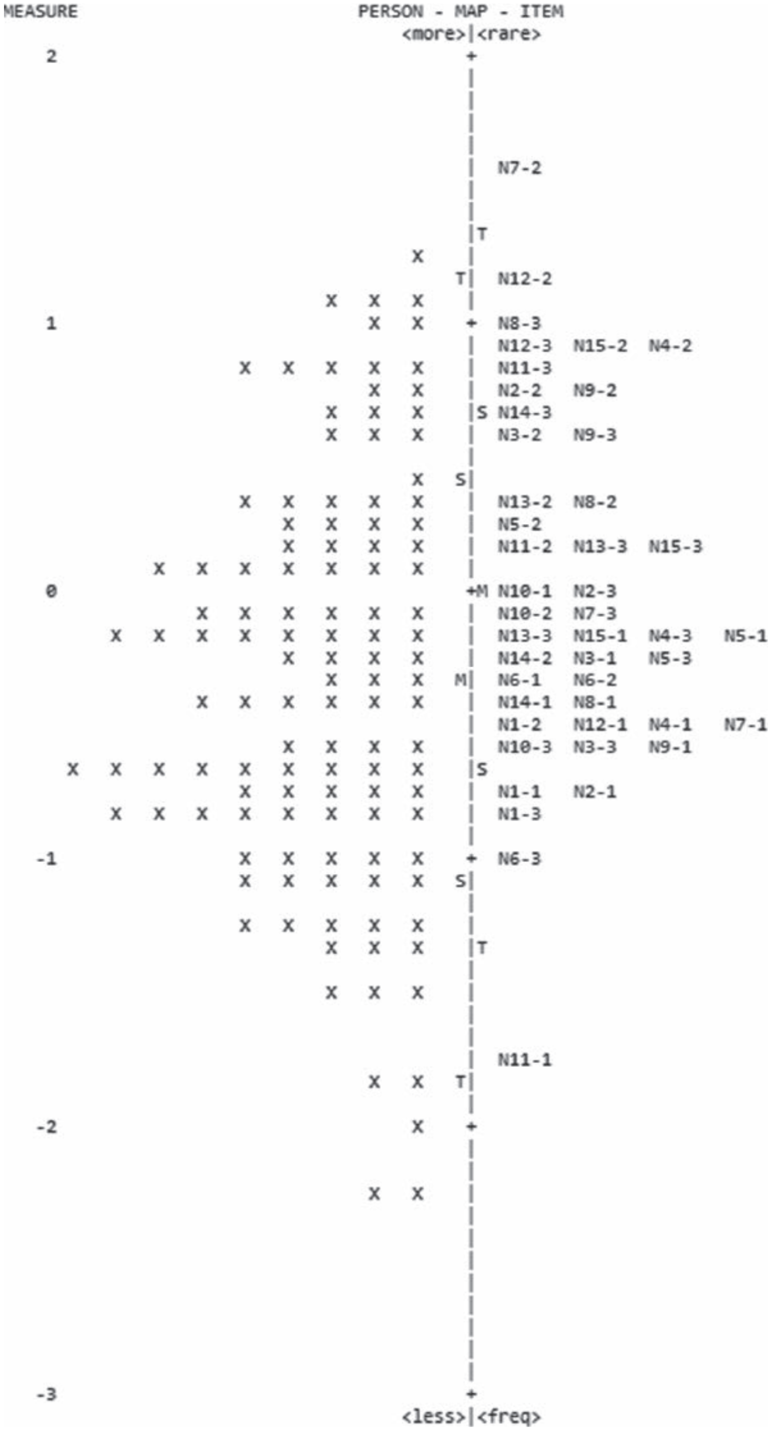
Table 4

Summary Table for Item Measures for the Results of the Lunic Numbers Subtest of the LLM

	Total Score	Measure	Real *SE*	Infit *MNSQ*	Infit *ZSTD*	Outfit *MNSQ*	Outfit *ZSTD*
M	51.30	.00	.21	1.00	-1.00	1.02	-1.00
P. SD	15.80	.66	.02	.09	1.30	.15	1.40
S. SD	16.0	.67	.02	.09	1.30	.15	1.40
Max	93.00	1.57	.28	1.26	3.10	1.53	3.80
Min	18.00	-1.77	.20	.82	-2.80	.77	-2.8
REAL RMSE	.21	*TRUE SD*	.62	SEP	2.98	ITEM REL	.90
MODEL RMSE	.21	*TRUE SD*	.63	SEP	3.04	ITEM REL	.90
SE OF PERSON MEAN			.10				

Note. SD = Standard deviation; P. SD = Population standard deviation; S. SD = Sample standard deviation; SE = Standard error; Max. = maximum value; Min = minimum value; MNSQ = mean-squared; ZSTD = Standardized z-scores; RSME = square-root of the average error variance; SEP = Separation; ITEM REL = Item reliability.

Figure 2
The Wright Map for Items and Persons for the Results of the Lunic Numbers Subtest of the LLM



```
MEASURE                          PERSON - MAP - ITEM
                                     <more>|<rare>
    2                                      +
                                           |
                                           |
                                           |
                                           |
                                           |  N7-2
                                           |
                                           |
                                           |T
                                      X    |
                                          T|  N12-2
                              X   X   X    |
    1                             X   X    +  N8-3
                                           |  N12-3  N15-2  N4-2
                      X   X   X   X   X    |  N11-3
                                  X   X    |  N2-2   N9-2
                              X   X   X    |S N14-3
                              X   X   X    |  N3-2   N9-3
                                           |
                                      X   S|
                      X   X   X   X   X    |  N13-2  N8-2
                          X   X   X   X    |  N5-2
                          X   X   X   X    |  N11-2  N13-3  N15-3
              X   X   X   X   X   X   X    |
    0                                      +M N10-1  N2-3
                  X   X   X   X   X   X    |  N10-2  N7-3
          X   X   X   X   X   X   X   X    |  N13-3  N15-1  N4-3   N5-1
                          X   X   X   X    |  N14-2  N3-1   N5-3
                              X   X   X   M|  N6-1   N6-2
                  X   X   X   X   X   X    |  N14-1  N8-1
                                           |  N1-2   N12-1  N4-1   N7-1
                          X   X   X   X    |  N10-3  N3-3   N9-1
      X   X   X   X   X   X   X   X   X    |S
                      X   X   X   X   X    |  N1-1   N2-1
          X   X   X   X   X   X   X   X    |  N1-3
                                           |
   -1                 X   X   X   X   X    +  N6-3
                      X   X   X   X   X   S|
                                           |
                      X   X   X   X   X    |
                              X   X   X    |T
                                           |
                              X   X   X    |
                                           |
                                           |
                                           |  N11-1
                                  X   X   T|
                                           |
   -2                                 X    +
                                           |
                                           |
                                  X   X    |
                                           |
                                           |
                                           |
                                           |
                                           |
                                           |
                                           |
                                           |
   -3                                      +
                                     <less>|<freq>
```

Analysis of the *Lunic* writing subtest. The results for the *Lunic* writing subtest were not as strong as for the *Lunic* numbers subtest. The results can be seen in Table 5. Person reliability was .77 when nine people were eliminated due to a ceiling effect. Person separation was 1.85, which indicated that the instrument could not separate the participants into strong, average, and poor phonetic coding ability groups. Item reliability was at a borderline acceptable coefficient of .73 The results are shown in Table 6. As can be seen in the Wright map (Figure 3), there is a considerable ceiling effect with six people answering all items perfectly and many participants having a higher logit than the most difficult-to-answer items. Furthermore, the item separation was 1.63, with most items clumped together between -0.5 and 0.5 logits. An analysis of the dimensionality of the instrument explained 23.5 percent of the variance with 6.14 eigenvalues. The first contrast had an eigenvalue of 1.86, explaining 7.1 percent of the variance. With such low eigenvalues, it is probable that this instrument is unidimensional and assesses the participants' phonetic coding abilities.

Table 5
Summary Table for Person Measures for the Results of the Lunic Writing Subtest of the LLM

	Total Score	Measure	Real *SE*	Infit *MNSQ*	Infit *ZSTD*	Outfit *MNSQ*	Outfit *ZSTD*
M	11.20	.35	.55	1.00	.10	.99	.10
P. SD	4.50	1.20	.14	.08	.60	.15	.70
S. SD	4.50	1.21	.14	.09	.60	.15	.70
Max	19.00	3.02	1.05	1.21	1.60	1.36	1.70
Min	2.00	-2.27	.46	.80	-2.20	.62	-2.10
REAL RMSE	.57	*TRUE SD*	1.06	SEP	1.85	PER REL	.77
MODEL RMSE	.57	*TRUE SD*	1.06	SEP	1.88	PER REL	.78
SE OF PERSON MEAN			.11				

Note. SD = Standard deviation; P. SD = Population standard deviation; S. SD = Sample standard deviation; SE = Standard error; Max. = maximum value; Min = minimum value; MNSQ = mean-squared; ZSTD = Standardized z-scores; RSME = square-root of the average error variance; SEP = Separation; PER REL. = Person reliability.

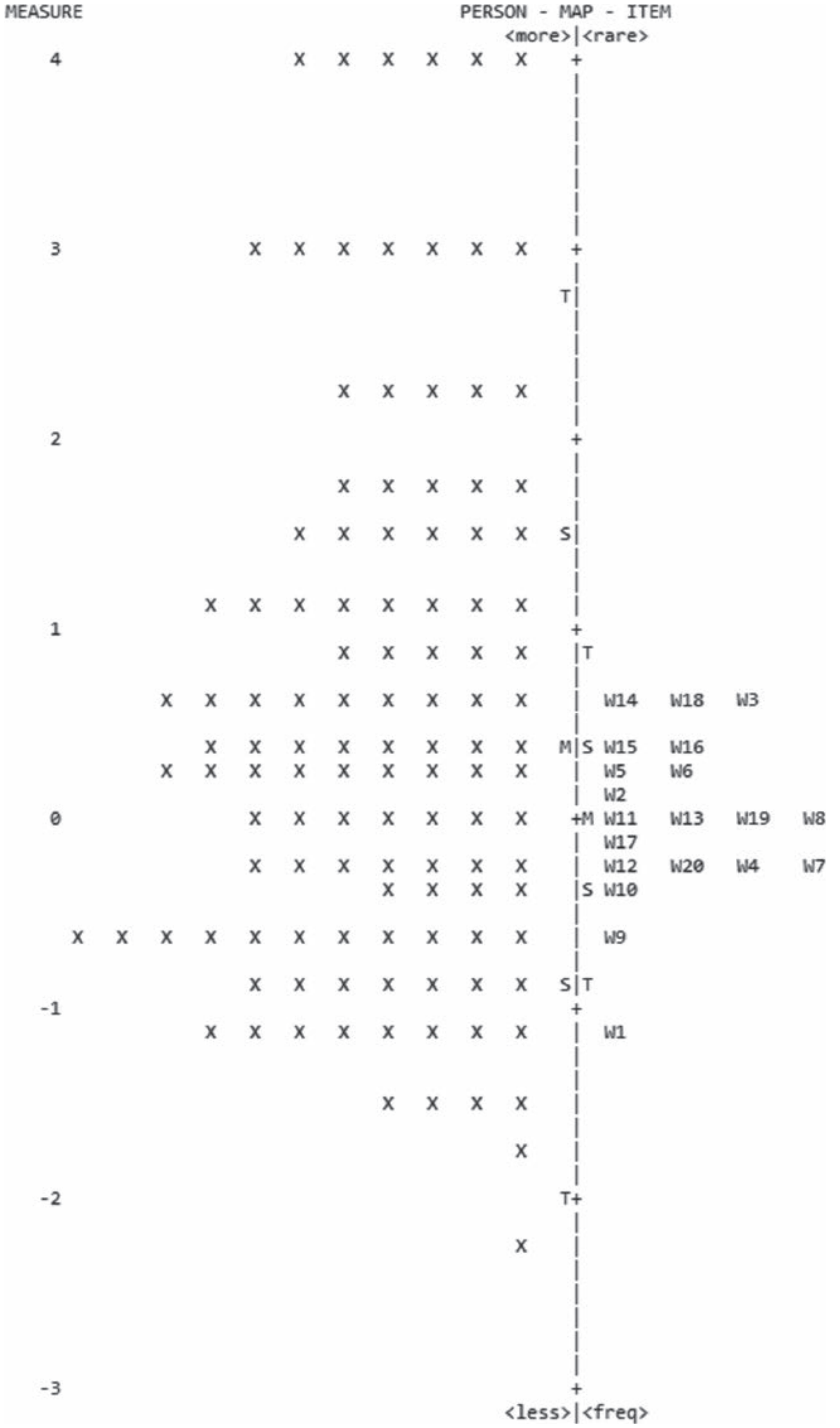
Table 6
Summary Table for Item Measures for the Results of the Lunic Writing Subtest of the LLM

	Total Score	Measure	Real *SE*	Infit *MNSQ*	Infit *ZSTD*	Outfit *MNSQ*	Outfit *ZSTD*
M	68.60	.00	.22	1.00	.00	.99	-1.00
P. SD	8.90	.43	.01	.14	1.50	.23	1.50
S. SD	9.20	.44	.01	.14	1.60	.23	1.50
Max	92.00	.63	.26	1.25	2.60	1.41	2.60
Min	55.00	-1.18	.21	.77	-2.80	.68	-2.60
REAL RMSE	.22	*TRUE SD*	.37	SEP	1.63	ITEM REL	.73
MODEL RMSE	.22	*TRUE SD*	.37	SEP	1.69	ITEM REL	.74
SE OF PERSON MEAN			.10				

Note. SD = Standard deviation; P. SD = Population standard deviation; S. SD = Sample standard deviation; SE = Standard error; Max. = maximum value; Min = minimum value; MNSQ = mean-squared; ZSTD = Standardized z-scores; RSME = square-root of the average error variance; SEP = Separation; ITEM REL = Item reliability.

Figure 3

The Wright Map for Items and Persons for the Results of the Lunic Writing Subtest of the LLM



```
MEASURE                                  PERSON - MAP - ITEM
                                              <more>|<rare>
    4                      X   X   X   X   X   X    +
                                                    |
                                                    |
                                                    |
                                                    |
                                                    |
                                                    |
                                                    |
    3                  X   X   X   X   X   X   X    +
                                                    |
                                                   T|
                                                    |
                                                    |
                                                    |
                               X   X   X   X   X    |
                                                    |
    2                                               +
                                                    |
                               X   X   X   X   X    |
                                                    |
                           X   X   X   X   X   X   S|
                                                    |
                                                    |
                   X   X   X   X   X   X   X   X    |
    1                                               +
                               X   X   X   X   X    |T
                                                    |
               X   X   X   X   X   X   X   X   X    |  W14   W18   W3
                                                    |
                   X   X   X   X   X   X   X   X   M|S W15   W16
               X   X   X   X   X   X   X   X   X    |  W5    W6
                                                    |  W2
    0                  X   X   X   X   X   X   X    +M W11   W13   W19   W8
                                                    |  W17
                       X   X   X   X   X   X   X    |  W12   W20   W4    W7
                                   X   X   X   X    |S W10
                                                    |
       X   X   X   X   X   X   X   X   X   X   X    |  W9
                                                    |
                       X   X   X   X   X   X   X   S|T
   -1                                               +
                   X   X   X   X   X   X   X   X    |  W1
                                                    |
                                                    |
                                   X   X   X   X    |
                                                    |
                                               X    |
                                                    |
   -2                                              T+
                                                    |
                                               X    |
                                                    |
                                                    |
                                                    |
                                                    |
                                                    |
   -3                                               +
                                              <less>|<freq>
```

Discussion and Conclusion

The purpose of this study was to pilot and validate the LLM as a reliable and valid instrument to measure participants' phonological awareness. The results of the Rasch analyses for the composite test were convincing to use as an aptitude test. However, as the subtests were designed to measure different components of language aptitude, it is necessary that both subtests are reliable and seem to measure what they are purported to measure. The reliability coefficients and spread of items for the *Lunic* numbers subtest suggest that it is a reliable instrument to use with this population. The item dimensionality indicated that there might be another construct within the subtest. One possibility is that participants are not only relying on phonemic discrimination and coding abilities, but they are also heavily reliant on their memory to attend to this test.

The *Lunic* writing subtest was of more concern. There was a clear ceiling effect with this subtest. Furthermore, the test had mediocre item reliability estimates and poor separation. Clearly there needs to be more difficult items for this subtest. One possible alteration to this subtest is to include more difficult items. The first ten items on this test consisted of two *Lunic* characters. Items 11 to 18 were comprised of three characters. However, the last two items on this test were of four characters. If the test had more items of such characters, there might be more spread on the Wright map. Another possible revision to the test to improve reliability is to change the training session. In the current version, there was a training session for each of the four sets. One suggestion to increase the difficulty is to have the participants learn the associated sounds at the beginning of the test, as in the *Lunic* numbers subtest, and then do all of the items. In future pilotings of this subtest, researchers should make these two alterations and determine if they improve the reliability of this instrument.

References

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Carroll, J. B., & Sapon, S. (1959). *The Modern Languages Aptitude Test.* Psychological Corporation.

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171. https://doi.org/10.1191/1362168805lr161oa

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*(3), 379–400. https://doi.org/10.1017/S0272263197003045

Kiss, C., & Nikolov, M. (2005). Developing, piloting, and validating an instrument to measure young learners' aptitude. *Language Learning*, *55*(1), 99–150. https://doi.org/10.1111/j.0023-8333.2005.00291.x

Linacre, J. M. (2009). Winsteps® RASCH Measurement [Computer software]. Version 4.0.0. Retrieved from http://www.winsteps.com/index.htm

Ortega, L. (2009). *Understanding second language acquisition*. Hodder Education.

Pimsleur, P. (1966). *The Pimsleur Language Aptitude Battery*. Harcourt, Brace, Jovanovic.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*(1), 45–99. https://doi.org/10.1111/0023-8333.21997002

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes, and learning conditions in second language acquisition. *Second Language Research*, *17*(4), 368–392. https://doi.org/10.1177/026765830101700405

Robinson, P. (2002a). Effects of individual differences in intelligence, aptitude, and working memory on adult incidental SLA: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–266). John Benjamins.

Robinson, P. (2002b). Learning conditions, aptitude complexes, and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211-266). John Benjamins.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. https://doi.org/10.1017/S0267190505000036

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In R. M. DeKeyser (Ed.), *Practice in a second language* (pp. 256–286). Cambridge University Press.

Ross, S., Yoshinaga, N., & Sasaki, M. (2002). Aptitude-exposure interaction effects on wh-movement violation detection by pre-and-post-critical period Japanese bilinguals. In P. Robinson (Ed.), *Language learning & language teaching* (pp. 267–299). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.2.14ros

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning*, *43*(3), 313–344. https://doi.org/10.1111/j.1467-1770.1993.tb00617.x

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses*. New York, NY: Peter Lang.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*(2), 255–283. https://doi.org/10.1002/j.1545-7249.2007.tb00059.x

Sick, J. R. (2007). *The learner's contribution: Individual differences in language learning in a Japanese high school* [Doctoral dissertation, Temple University]. https://digital.library.temple.edu/digital/

Sick, J. R., & Irie, K. (2001). Investigating the role of language aptitude in EFL courses in Japan. In P. Robinson, M. Sawyer, & S. Ross (Eds.), *Second language acquisition research in Japan* (Vol. 4, pp. 129-141). JALT Applied Materials Series.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 45–68). John Benjamins Publishing Company.

Snow, R. E. (1991). Aptitude–treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, *59*(2), 205–216. https://doi.org/10.1037/0022-006X.59.2.205

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.), *Individual differences and universals in langauge learning aptitude* (pp. 119–154). Newbury House Publishers.