The Korean Version of the Right Hemisphere Language Battery: Adaptation and Reliability

Jihyeon Yun, Jee-Hye Chung, Yeong-Wook Kim, Il-Young Jung

Department of Linguistics, Chungnam National University, Daejeon, Korea
Department of Rehabilitation Medicine, Chungnam National University Hospital, Daejeon, Korea
Department of Rehabilitation Medicine, Chungnam National University Sejong Hospital, Sejong, Korea

Correspondences: Il-Young Jung, MD, PhD and Yeong-Wook Kim, MD, PhD
Department of Rehabilitation Medicine, Chungnam National University Sejong Hospital, 20 Bodeum 7-ro, Sejong 30099, Korea
Tel: +82-44-995-4760
Fax: +82-44-995-5979
E-mail: {102onez, kyu0922}@cnuh.co.kr

Received: July 4, 2022
Revised: July 29, 2022
Accepted: August 1, 2022

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) to IJ (NRF-2021R1C1C1005188) and the Chungnam National University Sejong Hospital Research Fund 2021 to YK (2021-S2-005).

Objectives: This study presents the Korean version of the Right Hemisphere Language Battery (K-RHLB), reports the preliminary results in healthy subjects, and evaluates inter- and intra-rater reliability. Methods: Fifty-two healthy patients underwent the RHLB, which was adapted to the Korean language. The rater reliability of the Emphatic Stress Test, which is a subtest of the battery, was examined. For inter-rater reliability, three raters evaluated the focus prosody of 35 randomly chosen subjects. For intra-rater reliability, one rater re-evaluated the prosody within 2-week and 3-month intervals. Results: The total mean score of the K-RHLB result was 110.75 ± 2.90 out of 116. The mean percentages of correct responses in the subtests were high and appeared in the following order: Discourse Analysis (99.9%), Written Metaphor Test (99.6%), Lexical Semantic Test (99.3%), Picture Metaphor Test (98.7%), Emphatic Stress Test (94.4%), Inference Test (83.5%), and Humor Test (76.5%). In addition, the inter-rater agreement was high (Gwet's chance-corrected agreement coefficients AC1 of 0.83 ± 0.02, p < 0.001), as was the intra-rater consistency (Gwet's AC1 0.93 ± 0.01, p < 0.001). However, some changes in evaluation criteria were confirmed for retest intervals longer than 2 months [Cochran's Q (2) = 13.82, p < 0.001]. Conclusion: The preliminary results of the K-RHLB in healthy subjects showed overall high scores. Both inter- and intra-rater reliabilities were acceptable. The K-RHLB can be a starting point for more detailed assessment and elaborate intervention in patients with right hemisphere damage. The battery can also be used to evaluate neurodevelopmental disorders with symptoms such as deficits in understanding non-literal languages, semantic integration, and perception and production of appropriate prosody.

Keywords: Language tests, Cognition disorders, Social communication disorder, Speech disorders, Wit and humor

The left cerebral hemisphere is primarily responsible for language function. The phenomenon by which the unilateral cerebral hemisphere exerts more influence on language function is referred to as language lateralization (Knecht et al., 2000). The possibility of asymmetric functional roles of both hemispheres has been studied previously (Rasmussen & Milner, 1977), and various efforts have been made to measure these roles (Bradshaw, Thompson, Wilson, Bishop, & Woodhead, 2017).

Because of language laterality, several studies have been conducted targeting the left cerebral hemisphere in relation to language. However, only a few studies have been conducted to investigate the role of the right cerebral hemisphere in language function. Bryan (1988) noted that when an aphasia test was performed on subjects with right hemisphere damage (RHD), the results did not differ significantly compared with those of normal subjects. However, the same study reported that subtle language problems, such as those involving metaphorical comprehension and understanding of humor, appeared in subjects with right hemisphere in-
juries (Bryan, 1988). A language function assessment conducted on a split human brain with conditions involving transection of the corpus callosum revealed that the right hemisphere performs a language function different from that of the left hemisphere (Gazzaniga, 2000).

A stroke in the right cerebral hemisphere causes one or more types of language impairment (Dyukova, Glazman, Titova, Kriushev, & Gamaleya, 2010). Several studies have used various approaches to determine the role of the right hemisphere in language function. One test, the Right Hemisphere Language Battery (RHLB), was developed in British English to raise awareness about language and communication impairments in patients with RHD and evaluate these conditions qualitatively and quantitatively. Studies related to this test have also been conducted (Bryan, 1995; Bryan & Hale, 2001).

Ramsey and Blake (2020) reported the various standardized tests used by speech-language pathologists in the United States for the assessment and screening of cognitive-communication disorders associated with right hemisphere stroke. These tests are listed in Table 1 (adapted from Ramsey & Blake, 2020, p. 746).

Some of the tests in Table 1 were adapted to Korean. Table 2 shows a few such tests, as well as an original test, with their test domains.

Table 1. Standardized tests used by speech-language pathologists in the United States for the assessment of cognitive-communication disorders with right hemisphere stroke (adapted from Ramsey & Blake, 2020)

|-------|-----------------------------------|--------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------|---------------------------------------------------------------|

Among these, various versions of the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) or Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) are widely used as screening tools. For the thorough assessment of cognitive deficits, neuropsychological assessment batteries such as the Korean version of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD; Nasreddine et al., 2005) are used as screening tools. For the thorough assessment of cognitive deficits, neuropsychological assessment batteries such as the Korean version of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD; Nasreddine et al., 2005) are used as screening tools.

Table 2. Comparisons of tests for assessing symptoms related to RHD deficits

<table>
<thead>
<tr>
<th>Test domain</th>
<th>K-MMSE-K</th>
<th>MMSE-DS</th>
<th>MMSE-KC</th>
<th>MMSE-K</th>
<th>MoCA-K</th>
<th>MoCA</th>
<th>K-MoCA</th>
<th>CERAD-K(N)</th>
<th>SNSB-II</th>
<th>SNSB-C</th>
<th>SNSB-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration time</td>
<td>5-10 min</td>
<td>10-15 min</td>
<td>30-40 min</td>
<td>90-120 min (Shorter for SNSB-C/D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test domain</td>
<td>DDAAE (RH function)</td>
<td>Korean MEC</td>
<td>K-RHLB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facial expression</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visuospatial</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-literal language</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-awareness</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prosody</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexico-semantic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discourse</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pragmatic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reasoning</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humor</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration time</td>
<td>(60-90 min for the entire tests)</td>
<td>90 min</td>
<td>20-50 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MMSE=Mini Mental State Examination (Folstein et al., 1975); K-MMSE=the Korean MMSE (Kang, Na, & Hahn, 1997); MMSE-K=the Korean version of MMSE (Park & Kwon, 1990); MMSE-KC=the Korean version of MMSE in the CERAD-K(N) (Lee et al., 2002); MMSE-DS=the Korean version of MMSE for Dementia Screening (Kim et al., 2010); MoCA=Montreal Cognitive Assessment (Nasreddine et al., 2005); K-MoCA=Korean-Montreal Cognitive Assessment (Kang, Park, & Lee, 2009); CERAD-K(N)=CERAD-K(N)=K Consortium to Establish a Registry for Alzheimer’s Disease neuropsychological assessment battery; SNSB=Seoul Neuropsychological Screening Battery (Kang & Na, 2003); SNSB-II=Korean MoCA-2nd edition (Kang, Jahng, & Na, 2012); SNSB-C=Korean MoCA-Core (Jahng, Na, & Kang, 2015); SNSB-D=Korean MoCA-Dementia Version (Ahn et al., 2010); DDAAE=Daegu Diagnostic Aphasia Examination (Jeong, 2008); RH=right hemisphere; Korean MEC=the Korean Montreal Protocol for Evaluation of Communication (Woo & Kim, 2019; Joanette et al., 2004); K-RHLB=the Korean version of the Right Hemisphere Language Battery (this study).
Jihyeon Yun, et al. • Adaptation and Reliability of the K-RHLB

er’s Disease neuropsychological assessment battery [CERAD-K(N); Lee et al., 2002] or Seoul Neuropsychological Screening Battery (SNSB; Kang & Na, 2003) can be used. Although comprehensive and intensive, these batteries mainly focus on assessing symptoms of dementia. The Daegu Diagnostic Aphasia Examination (DDAE; Jeong, 2006), specifically the section for right hemisphere brain function, and the Korean Montreal Protocol for Evaluation of Communication (Korean MEC protocol; Woo & Kim, 2019) are more language- and communication-oriented tests that are associated with RHD traits. Additionally, Ahn (2008) reported using the Mini Inventory of Right Brain Injury (MIRBI; Pimental & Kinsbury, 1989), translated to Korean by the author, for participant screening in the study which pertained to comprehension of Korean proverbs in RHD patients. Furthermore, Yang and Van Lancker Sidtis (2016) reported test results in RHD patients using the Korean version of the MIRBI.

The evaluation tools for language function should be developed using the country’s native language because language is heavily influenced by sociocultural factors. However, in South Korea, there is currently no well-established test for right cerebral hemisphere language function. Therefore, for the quantitative evaluation of right hemisphere language function, the current study reports the adaptation protocol of the Korean version of the Right Hemisphere Language Battery (K-RHLB), the preliminary test results for each subtest, and the reliability of the K-RHLB.

METHODS

Test Materials

The RHLB (Bryan, 1995) comprises seven subtests related to the symptoms reported for patients with RHD, particularly those for deficits in comprehension of metaphor (Winner & Gardner, 1977), semantic integration (Bryan, 1988; Caramazza, Gordon, Zurif, & DeLuca, 1976; Wapner, Hamby, & Gardner, 1981), appreciation of humor (Brownell, Michel, Powelson, & Gardner, 1983; Gardner, Ling, Flamm, & Silverman, 1975), lexical-semantic association (Coughlan & Warrington, 1978; Gainotti, Caltagirone, & Miceli, 1983; Joanette, Goulet, Hannequin, & Boeglin, 1990; Taylor & Warrington, 1973), and understanding or expressing intention based on situation and context (Bryan, 1988; Hirst, LeDoux, & Stein, 1984; Weintraub & Mesulam, 1983).

The original test materials were first translated into Korean and retained whenever possible. However, most non-literal phrases have been replaced with appropriate Korean expressions. Following the original battery, the K-RHLB (version 1.0) consists of seven subtests, six of which are formal tests (Picture Metaphor Test [PMT], Written Metaphor Test [WMT], Inference Test [IT], Humor Test [HT], Lexical Semantic Test [LST], and Emphatic Stress Test [EST]). The remaining subtest is a post-session evaluation (Discourse Analysis [DA]). Each formal test includes a practice item. The structures of the subtests and scoring system are analogous to the original RHLB (Bryan, 1995) and the Italian version of the RHLB (Batteria del Linguaggio dell’Emisfero Destro; Zanini, Bryan, De Luca, & Bava, 2005), except that the K-RHLB has a revision in the EST.

The EST evaluates prosody production by inducing emphatic stress on new information that contrasts with given information, where the stress can be delivered as changes in pitch, rhythm, and/or loudness. The K-RHLB (version 1.0) subdivided EST into the Emphatic Stress Test-Ga (EST-Ga) and Emphatic Stress Test-Na (EST-Na), with each comprising one practice item and five main items.

The EST-Ga was revised, as motivated by Lee and Park (2019), to include the induction of narrow-focus prosody using corrective focus. Narrow-focus prominence can be more salient than broad-focus prominence, which helps the examiner perceive and judge emphatic stress. For each item, a picture is presented, and the examiner asks the participant a simple yes-or-no question, emphasizing the target phrase that is always false (e.g., “Does the woman go to buy meat?”). The participant responds by correcting the target (e.g., “No, the woman goes to buy vegetables.”).

The EST-Na maintains the original form of the Production of Emphatic Stress (Bryan, 1995), wherein a set of two pictures with contrasting information is referred to for each item. The examiner points to the first picture and says the first half of the sentence; the subject then completes the second half, referring to the second picture. The maximum score is 5 for each section of the EST (Ga and Na), thus rendering a maximum of 10 points for the entire subtest.

Participants

A total of 52 native Korean speakers without language disorders participated in the study (mean age: 31.4 ± 7.6 years; 38 females; 48
right-handed, 1 left-handed, 3 ambidextrous). All the participants had normal or corrected-to-normal vision. Written informed consent was obtained from all the participants, and financial compensation was provided.

**Test Procedures**

This study was approved by the Institutional Review Board (CNUSH 2021-04-002) and was conducted in June and July 2021. Both the examiner and the participant wore face masks during the session to prevent infectious disease transmission. Each participant was individually tested in a quiet room by an examiner trained in linguistics and phonetics (first author). Figure 1 illustrates the test procedures performed. After providing instructions and filling out the demographic questionnaire, the K-RHLB (version 1.0) was conducted. In this study, the battery was performed in a fixed sequence, specifically in the order of PMT, WMT, IT, HT, LST, and EST. The test was conducted with the test materials (pictures or text) being presented to the participant via a computer monitor. Feedback was provided only for the practice items at the beginning of each formal subtest. Each session took around 25–60 minutes (approximately 40 minutes on average), depending on the pace of the participant. The DA was performed by the examiner after the session.

**Reliability Tests**

To verify rater reliability, the EST was chosen from the formal subtests as it is most likely to be affected by examiners’ subjective judgment criteria due to the gradient and less categorical nature of prosody. As for the other five formal tests, four (PMT, WMT, HT, and LST) were multiple-choice tasks, thus being free from rater degree-of-freedom, and the IT was evaluated based on the preset range of responses. Thus, the EST was subjected to the estimation of inter- and intra-rater reliability.

The evaluation scores for the EST were obtained from three raters, namely the examiner who conducted the audio-recorded sessions (Rater 1) and two additional raters who were qualified speech-language pathologists with more than 10 years of experience (Raters 2 and 3). The latter two raters listened to the recordings of 35 randomly chosen subjects and evaluated the emphatic stress on each item by assigning a score of 1 (appropriate focus prosody) or 0 (no prosodic saliency on/around the target).

Test-retest reliability was investigated based on Rater 1’s repetitive evaluations. The first session (Session 1) was the original test session conducted in person by Rater 1, as described above. Rater 1 re-evaluated the EST based on audio recordings 1–2 weeks (Session 2) and 2–3 months (Session 3) after the first session.

**Data Analysis**

A total of 4,316 data points (52 subjects × 83 evaluation entries) were obtained. All statistical analyses were performed using R software (R Core Team, Vienna, Austria; https://www.R-project.org/). The scores were summarized for each subtest, and the mean percentages of correct responses are provided.

In analyzing rater reliabilities in the EST, each data point was coded 1 or 0 for appropriate or inappropriate emphatic stress, respectively. Gwet’s chance-corrected agreement coefficients AC1
Adaptation and Reliability of the K-RHLB (Gwet, 2008) were calculated for inter-rater reliability using the `gwet.ac1.raw()` function in the R package `irrCAC`. A total of 1,050 data points (3 raters × 35 speakers × 10 items) were used.

To evaluate intra-rater reliability, McNemar’s Chi-square tests of symmetry were performed to assess changes within the rater’s scores using the `pairwiseMcnemar()` function of the `rcompanion` package in R with the “fdr” option for repeated-measure correction using the Benjamini and Hochberg method (Benjamini & Hochberg, 1995). Along with Gwet’s AC1, the overall intra-rater variability across the three sessions was estimated using Cochran’s Q using the `cochran.qtest()` function in the R package `RVAideMemoire`. The number of data points used was 1,050 (3 sessions × 35 speakers × 10 items).

RESULTS

Descriptive Statistics

The total mean score of the K-RHLB in 52 healthy subjects was 110.75 ± 2.90 out of 116 points (mean percentage correct: 95.5 ± 2.5%). Table 3 shows the mean scores along with the proportions of correct responses for each subtest. The DA, which is a post-session assessment, revealed that the highest mean score reached 99.9% of the maximum score. Among the six formal subtests, the percentage of correct responses was highest for the WMT (99.6%), followed by the LST (99.3%), PMT (98.7%), and EST (94.4%). The IT obtained a relatively low percentage of correct responses (83.5%), and the HT had the lowest percentage of correct responses (76.5%).

Figure 2 shows the distribution of the participants’ scores in terms of the proportion of correct responses. The differences in their overall ranges and interquartile ranges of scores across subtests indicate that subject variability is greater in the IT, the HT, and the EST than in the other four subtests.

Inter-rater Reliability

The number of cases with correct placement of emphatic stress was 328 (93.7%), 293 (83.7%), and 338 (96.6%) according to the judgments of Raters 1, 2, and 3, respectively. The contingency table for the three raters (correct response: 1; incorrect response: 0) is shown in Table 4. The raw percentage agreement was 79.1% between the three raters, with 350 cases each (10 items × 35 subjects). Gwet’s AC1 was 0.83 ± 0.02 (95% CI: 0.80-0.87, p < 0.001). Figure 3 compares the scores given by the three raters to each subject in the EST. The points given by Rater 1 ranged from 7 to 10 per subject, with 10 points being the most common and the counts decreasing towards 7 points. The points per subject given by Rater 2 also ranged from 7 to 10, but a score of 7 points was the most common, followed by 8 and 10. Rater 3 evaluated the subjects’ performance at 9 or 10 points. None of the participants gained a score lower than 7 from the three raters.

Table 3. Maximum score, mean score, standard deviation, and percentage of correct responses in 52 healthy subjects for each subtest of the K-RHLB

<table>
<thead>
<tr>
<th>Test (maximum score)</th>
<th>Score (mean± SD)</th>
<th>Percent correct (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture Metaphor (10)</td>
<td>9.87 ± 0.35</td>
<td>98.7</td>
</tr>
<tr>
<td>Written Metaphor (10)</td>
<td>9.96 ± 0.19</td>
<td>99.6</td>
</tr>
<tr>
<td>Inference (12)</td>
<td>10.02 ± 1.41</td>
<td>83.5</td>
</tr>
<tr>
<td>Humor (10)</td>
<td>7.65 ± 1.62</td>
<td>76.5</td>
</tr>
<tr>
<td>Lexical Semantic (20)</td>
<td>19.97 ± 0.40</td>
<td>99.3</td>
</tr>
<tr>
<td>Emphatic Stress (10)</td>
<td>9.44 ± 0.83</td>
<td>94.4</td>
</tr>
<tr>
<td>Discourse Analysis (44)</td>
<td>43.94 ± 0.24</td>
<td>99.9</td>
</tr>
</tbody>
</table>

Table 4. Contingency table of the evaluations of the emphatic stress test by three raters. Unit: count (percentage in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Rater 3: Incorrect</th>
<th>Rater 3: Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>Rater 2: Incorrect</td>
<td>Rater 2: Correct</td>
</tr>
<tr>
<td>Incorrect</td>
<td>3 (0.9%)</td>
<td>1 (0.3%)</td>
</tr>
<tr>
<td></td>
<td>1 (0.3%)</td>
<td>5 (1.4%)</td>
</tr>
<tr>
<td>Correct</td>
<td>3 (0.9%)</td>
<td>5 (1.4%)</td>
</tr>
<tr>
<td></td>
<td>1 (0.3%)</td>
<td>48 (13.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>274 (78.2%)</td>
</tr>
</tbody>
</table>

K-RHLB = Korean version of the Right Hemisphere Language Battery; SD = standard deviation.
Intra-rater Agreements

The contingency table of evaluations for the three sessions by Rater 1 is shown in Table 5 (correct response: 1, incorrect response: 0). The raw percentage of agreement between sessions was 96.6% for Sessions 1 and 2, 92.6% for Sessions 1 and 3, and 92.0% for Sessions 2 and 3. The raw intra-rater agreement across the three sessions was 90.6%. Gwet’s AC1 within the rater for the three sessions was 0.93 ± 0.01 (95% CI: 0.90-0.95, p < 0.001).

The results of the McNemar tests showed that the evaluation results from Session 1 and 2 were similar \( \chi^2(1) = 0.33, p = 0.56 \), whereas those of Session 3 significantly differed from those of both Sessions 1 \( \chi^2(1) = 7.54, p = 0.009 \) and 2 \( \chi^2(1) = 9.14, p = 0.008 \). Overall, the evaluation scores varied within the rater across repeated evaluations, reflecting the differences between the earlier two sessions and the last session [Cochran’s Q(2) = 13.82, p < 0.001].

DISCUSSION

Emphatic Stress and Rater Reliability

The EST was retained in the K-RHLB as it differentiated the RHD group from both the left hemisphere damage (LHD) and control groups, neither of which differed significantly from each other in terms of the Production of Emphatic Stress in the original battery [p. 21, Table 4.1 and p. 22, Table 4.2 in Bryan (1995)]. Although the results were lower than the scores obtained by neurotypical English subjects in Bryan (1995), the participants in this study obtained relatively high mean scores on the EST.

The current study provided novel results on the rater reliability of prosodic evaluation, wherein other factors were constant. The inter-rater reliability among the three raters was high, especially considering the difficulty of judging prosodic features in Korean (e.g., Lee & Cho, 2020). This positive result may be attributed to the fact that positive response rates were high across three raters (approximately 83–97%). Although not comparable in terms of study design, a previous report on rater reliabilities using the original RHLB (Bryan, 1995) similarly found a high concordance between two judges (Pearson correlation coefficient \( R = 0.89, p < 0.001 \)) in the DA of RHD and LHD subjects.

For intra-rater variability, the change in scores between two consecutive sessions held within 2 weeks of one another was trivial. In contrast, a statistically significant difference was observed between the second and third sessions. The total number of points given by the rater was 328 in Session 1, 330 in Session 2, and 314 in Session 3, with a decrease of 16 points noted from Sessions 2 to 3. It is important to note that the retest sessions (Sessions 2 and 3) were conducted based on the audio recordings. Since the study was in progress over the interval between the second and third evaluations, it is possible that the rater modified the judgment criteria. This result implies that an examiner’s judgment can vary with time and experience. Nevertheless, the assessment of emphatic stress was relatively stable in healthy subjects with good intra-rater consistency. A previous study on the original RHLB provided another example of test-retest reliability (Bryan, 1995), where the RHD and LHD groups were tested 1–6 weeks post-onset and re-examined 14–20 weeks post-onset. It yielded consistent results between the two sessions, except for significant differences in the LST and DA for the RHD group and the IT, LST, and EST for the LHD group.
Overall Test Results

The test scores of healthy subjects were high across the board, with noticeable individual variations in the IT and HT. The subjects obtained remarkably high mean scores in the two types of metaphor tests (PMT and WMT) and the LST. However, assessing for the IT was not without difficulties as subjects had more freedom in the manner and length of their responses, despite the preset criteria for evaluation. The HT was the most difficult among the formal tests. An explanation for this result might be that a multiple-choice test may not be straightforward enough to find the “correct” punchline of a humorous story. However, it does not explain the fact that the score for the Appreciation of Humor is much higher in Bryan (1995) with a mean score of 9.70 ± 0.47 in 30 neurologically normal subjects. Alternatively, the current result may be related, in part, to the fact that there were subjects who were not familiar with the type of jokes with a punchline, which involves the transition of interpretation and development against the initial set-up of the story.

The scores of the DA were very high, the mean of which reached nearly 100% correct responses. This result can be due to some confounding factors, such as lack of time for natural communication between the examiner and subject during the session and differences in sociocultural norms and expectations. For example, it is common not to use humor or jokes in the Korean culture to indicate sincerity. With this tendency to be “in the normal range,” most of the discourses would be accepted as typical in terms of humor (which is an evaluation criterion in the DA that is independent of the HT), whereas the discourse would be marked down for “offensive” humor (e.g., sarcasm or irony). Further discussion requires test data in patients with “Pragmatic Aphasia” (Joanette & Ansaldo, 1999), or “Apragmatism” (Myers, 2001), i.e., the symptoms associated with RHD.

The RHLB in Clinical Use

Aphasia is a typical symptom of a lesion that occurs in the left hemisphere, and the types of aphasia can be roughly classified. Therefore, several studies related to language function in the left hemisphere have been conducted. However, the language function of the right hemisphere is less obvious than that of the left hemisphere. In addition, since the right hemisphere is also related to cognitive function, its significance in language function has been clinically overlooked.

Communication problems caused by RHD are not limited to language issues. Patients with RHD display signs of cognitive decline, such as neglect syndrome, as well as deficits in concentration, memory, and executive ability (Myers, 1999; Tompkins, Klopouchniotou, & Scott, 2013). It also causes problems in abstract language comprehension, such as metaphors, inference making, and understanding jokes. Patients with RHD show deficits in prosodic structures and have difficulty communicating using nonverbal elements (Baum & Dwivedi, 2003). These problems negatively affect the activities of daily living, social participation, and quality of life of the patients (Blake, 2006; Lehman & Tompkins, 2000).

According to a recent study, the recovery of aphasia after a stroke occurs not only in the left cerebral hemisphere but also in the right cerebral hemisphere through complex interrelationships (Li, Mukadam, & Kiran, 2022; Weiller et al., 1995). Each study has multiple heterogeneous results; however, it seems certain that the functional activity of the left and right cerebral hemispheres changes dynamically during recovery from aphasia (Hartwigsen & Saur, 2019). This means that the language function of the right cerebral hemisphere greatly affects the recovery from aphasia caused by lesions in the left cerebral hemisphere. That is, the preservation of the speech function of the right cerebral hemisphere can positively affect the patient’s recovery from aphasia.

Therefore, it is important to acknowledge the clinical significance of RHLB in patients with stroke. Detecting and evaluating these symptoms in the initial stages of right cerebral hemisphere injury is very important for establishing a treatment plan for the patient and providing early interventional treatment. In addition, it can provide vital information for understanding the recovery mechanism of aphasia caused by left brain lesions and predicting the prognosis.

A limitation of the present version of the K-RHLB is that, in principle, it does not impose any time limits on the subject’s response. Because the deficits in RHD are typically related to delays or inefficiencies rather than a total failure of the processes, Tompkins (2012) suggested emphasizing the speed of response in the treatment of the RHD population. As the current study is based on a healthy population, RHD and LHD patients should be exam-
This study reviewed the adaptation of the K-RHLB (version 1.0) for assessing language communication deficits and provides preliminary test results in healthy subjects, along with inter- and intra-rater reliabilities for the evaluation of emphatic stress. The mean scores were very high in metaphor comprehension and picture-word matching tasks, whereas they were relatively low in the integration of information and completion of jokes.

Pragmatic language and communication skills require the incorporation of linguistic knowledge and an understanding of interlocutors’ intentions and social norms in a particular context. This complexity complicates the quantification and evaluation of RHD syndromes. The K-RHLB can be a starting point for more detailed assessment and elaborate intervention in patients with RHD. In addition, the K-RHLB can be used to evaluate individuals with neurodevelopmental disorders with symptoms such as deficits in understanding non-literal languages, semantic integration, and perception and production of appropriate prosody [e.g., for schizophrenia (Pawęłczyk, Kotlicka-Antczak, Łojek, Ruszpeł, & Pawęłczyk, 2018)]. The normative scores of the K-RHLB for healthy subjects and patients with right and left hemisphere damage should be established in future studies.

ACKNOWLEDGMENTS

We would like to thank Karen Bryan for her permission to adapt the RHLB to the Korean language. This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) to IJ (NRF-2021R1C1C1005188) and the Chungnam National University Sejong Hospital Research Fund 2021 to YK (2021-S2-005).

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare, and there is no financial interest to report.


Gazzaniga, M. S. (2000). *Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition?*. *Brain, 123*(7), 1293-1326.


국문초록

한국어판 우뇌언어기능평가: 번안과 신뢰성 검증

윤지현1, 정지혜2, 김영욱3, 정일영2,3

1 충남대학교 언어학과, 2 충남대학교병원 재활의학과, 3 세종충남대학교병원 재활의학과

배경 및 목적: 본 연구에서는 우뇌언어기능평가를 번안하여 정상 성인의 검사 결과를 보고하고, 검사자간, 검사자 내 신뢰도를 검증하고자 하였다. 방법: 우뇌언어기능평가를 한국어로 번안하고 건강한 성인 52명을 대상으로 검사하였다. 하위 검사의 각 결과를 바탕으로 분석하였다. 검사자간 신뢰도는 3명 검사자의 평가 결과를 바탕으로 분석하였다. 검사자 내 신뢰도는 3명 검사자의 평가 결과를 바탕으로 분석하였다. 검사자간 신뢰도는 3명 검사자의 평가 결과를 바탕으로 분석하였다. 검사자 간 신뢰도는 3명 검사자의 평가 결과를 바탕으로 분석하였다.

방법: 우뇌언어기능평가를 한국어로 번안하고 건강한 성인 52명을 대상으로 검사하였다. 하위 검사인 강조강세검사의 결과를 토대로 신뢰도 분석을 위해 무작위로 채택한 35명의 발화 녹음 자료를 사용하였다. 검사자 간 신뢰도는 3명의 평가 결과를 바탕으로 분석하였다. 검사자 내 신뢰도는 3명의 평가 결과를 바탕으로 분석하였다. 검사자 간 신뢰도는 3명의 평가 결과를 바탕으로 분석하였다. 검사자 내 신뢰도는 3명의 평가 결과를 바탕으로 분석하였다. 검사자 간 신뢰도는 3명의 평가 결과를 바탕으로 분석하였다.

결과: 한국어판 우뇌언어기능평가 결과 전체 평균은 110.75 ± 2.90점 (116점 만점)이었다. 하위검사별 정반응 백분율은 담화분석 99.9%, 은유글검사 99.6%, 어휘의미검사 99.3%, 은유그림검사 98.7%, 강조강세검사 94.4%, 추론검사 83.5%, 유머검사 76.5%었다. 검사자간 신뢰도는 Gwet AC1 계량은 0.83 ± 0.02 (p < 0.001)로 높게 나타났다. 검사자간 신뢰도는 Gwet AC1 계량은 0.83 ± 0.02 (p < 0.001)로 높게 나타났다. 검사자간 신뢰도는 Gwet AC1 계량은 0.83 ± 0.02 (p < 0.001)로 높게 나타났다. 검사자간 신뢰도는 Gwet AC1 계량은 0.83 ± 0.02 (p < 0.001)로 높게 나타났다. 검사자간 신뢰도는 Gwet AC1 계량은 0.83 ± 0.02 (p < 0.001)로 높게 나타났다.

논의 및 결론: K-RHLB 검사에서 정상 성인은 전반적으로 높은 점수를 보였고, 강조강세산출의 검사자간, 검사자내 신뢰도는 용인 가능한 수준으로 나타났다.

핵심어: 언어검사, 인지장애, 사회의사소통장애, 말장애, 유머

이 연구는 세종충남대학교병원 연구기금(2021-S2-005)과 대한민국 정부로부터 자금을 지원받는 한국연구재단(NRF-2021R1C1C1005188)의 지원으로 수행되었음.

참고문헌


ORCID

윤지현(제1저자, 강사, https://orcid.org/0000-0003-3251-7672); 정지혜(공동저자, 언어치료사 https://orcid.org/0000-0002-3953-6553); 김영욱(공동교신저자, 교수 https://orcid.org/0000-0003-3282-1164); 정일영(공동교신저자, 교수 https://orcid.org/0000-0001-8204-8195)