

Feasibility of a novice electronic psychometric assessment system for cognitively impaired

Haemi Jee¹, Jaehyun Park^{2,*}

¹Department of Physical Therapy, Namseoul University, Cheonan, Korea

²Department of Information and Communication, Inha University, Incheon, Korea

The assessment and rehabilitation of patients with cognitive dysfunction is a field that currently requires assistive technology. While the paper-and-pencil test, such as the line tracing test, is one of the commonly used assessment methods for cognitive dysfunction, accuracy, and time-consuming assessment process needed technological application. The aim of this study was therefore to establish a computer-based real-time assessment system (e-system) for patients without compromising the usefulness of the conventional paper-and-pencil based user tools with 50 healthy participants. The comparison of the e-system with the golden-standard assessment (evaluator) results showed high concordance correlation coefficients of 0.89 and 0.87 and small effective sizes of 0.27 and 0.27 between two repeated measures. The Bland-Alt-

man plots also showed smaller degree of error and greater repeatability in comparison to the repeated measures. Moreover, the accuracy rates of 96.5% and 96.4% were shown. The results indicated feasibility of the novice e-system. The e-system may assist rehabilitation specialists to assess and diagnose patients with cognitive dysfunction. This system can be applied to a range of assessment and rehabilitation modalities based on pen and paper. It can also be used for various patients such as those with Parkinson disease, stroke, or different forms of brain lesions.

Keywords: Cognitive dysfunction, Golden standard, Psychometric test, Line tracing test


INTRODUCTION

Cognitive dysfunction that may be caused by brain damage or dysfunction is recognized by conducting neurocognitive tests such as psychometric tests or neuro-physiologic tests along with general clinical diagnosis. These test methods are useful in the clinical environment by detecting cognitive dysfunction early through the detection of adverse reactions even when no clinical symptoms are seen (Jeong et al., 2017).

Since cognitive dysfunction can appear in various symptoms, various types of tests are organized in the form of a comprehensive test rather than one test method to examine various functional states of patients. For example, in the case of the line bisection test (LBT), the degree of unilateral ignorance is evaluated by measuring the symptom of ignoring one of the left and right directions, and the judgment impairment that cannot be found in order from

1 to 10 is evaluated (Ferber and Karnath, 2001; Weissenborn et al., 2001). In the case of the line tracing test (LTT), the motor speed and accuracy are evaluated, and the Number Connection Test-A (NCT-A) is the mental-motor speed (psychomotor speed), visual scanning efficiency, sequencing, attention, concentration, etc. are evaluated, and the Number Connection Test-B (NCT-B) is attention set shifting ability, visual scanning efficiency, continuity, attention, and concentration are evaluated (Kerai et al., 2012; Li et al., 2013; Luo et al., 2019). Such psychometric test tools such as LTT, NCT-A, and NCT-B have low sensitivity and can be easily used. For such reasons, these tools are being widely used due to the advantage of being applicable to patients without clinical abnormality (Li et al., 2013).

The psychometric tests in general can be broadly divided into paper-and-pen or computer-based test methods (Luo et al., 2019). Paper-and-pen methods have the advantage of being a test tool

*Corresponding author: Jaehyun Park  <https://orcid.org/0000-0001-5699-9034>
Department of Information and Communication, Inha University, 100 Inha-ro,
Michuhol-gu, Incheon 22212, Korea
E-mail: jhyun@inha.ac.kr

Received: November 1, 2020 / Accepted: December 6, 2020

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

that can be easily carried out without being limited by the place and situation. The basic method of testing tools based on paper-and-pen is to print specific patterns on plain paper and have the patient draw a line using a pen in accordance with the regulations informed by the evaluator. In the case of the LTT test mentioned above, patients simply draw a line from the start point to the end point without taking off the pen without deviating from the given line. In the case of NCT-A or NCT-B, this is a test that sequentially connects patterns with numbers printed to the patient (Weissenborn et al., 2001). Despite the simplicity, the tests measure the degree of tremor of the arm, cognitive ability, movement accuracy to assess initial condition or rehabilitation outcome. Patients with real cognitive impairments show problems such as drawing outside the line or connecting incorrect numbers (Li et al., 2013). At the end of the test, an evaluation is made to determine the extent of the present symptoms based on the results.

As for the existing pen-and-paper based methods, the evaluation is also being carried out manually. Such manual evaluation method not only consumes the time and effort of a medical professional, but also makes it possible to generate an evaluator error (de Joode et al., 2010; Jee et al., 2015). In particular, in the case of the LTT test, it takes a lot of time because it is necessary to divide the given path into 365 pieces and determine the line of the corresponding area (Rossetti et al., 2016). Such test method also leads to reduced accurate evaluation results to the patient as error may occur during the repeated measurements or between evaluators (Jee et al., 2015). Moreover, since the test results must be preserved as medical records, there is also the hassle of scanning and storing test results. Therefore, there is a need for a new type of diagnosis system to maintain the existing evaluation method using a pen and paper, shorten the evaluation time, and guarantee the accuracy of evaluation results (de Joode et al., 2010; Ferber and Karnath, 2001).

Currently, computer-based examinations have limitations in providing patients with a changed measurement environment because they propose computer monitor measurement methods from paper-and-pen-based measurement methods (Jee et al., 2015). In order to solve this problem, this study aimed to provide a more precise and efficient system by implementing a system capable of real-time examination and evaluation while maintaining the same paper-and-pen based examination methods currently used for examination and diagnosis in hospitals.

This study examined the feasibility of a smart-type LTT system (e-system) made for assessing cognitively impaired. The LTT evaluation results by the e-system were compared with the golden standard assessment method.

MATERIALS AND METHODS

Participants

In order to evaluate the effectiveness and performance of the electronic pen (e-pen) based LTT system proposed in this paper, a test was conducted with 50 healthy adult participants. The criteria for selection of subjects were conducted by selecting 20 healthy adult men and 30 women between the ages of 19 and 32. The average age and educational background of the participants were 22.8 ± 2.57 years and 14.9 ± 1.70 years. Prior to conducting this study, the purpose and method of this study were explained in detail, and the experiment was conducted after confirming participation in this study through oral and signed signatures. This experimental process was approved by the Research Ethics Committee (170227-1A).

To confirm the cognitive function of the healthy adult subjects, the Korean version-Mini Mental State Examination (K-MMSE) was conducted, and as a result, the average K-MMSE score of the subjects was 29.4 ± 0.82 (Kang et al., 2016). The LTT test duration (sec) by each subject and the evaluation duration (sec) of the evaluator and the computer for the LTT results along with the accuracy rate (%) were also measured.

The LTT test results were scored in two ways. First, a computer program automatically scoring, and the scoring time for the test results was made in real time and was less than 1 second. In order to accurately compare with the existing scoring method, the evaluator directly scored using a pen and ruler.

Novice electronic psychometric assessment system: e-system

This study used custom-made electronic pen-based evaluation and rehabilitation system consists of an electronic pen, a smart paper printed with location information, and a PC-based evaluation software. When a subject touches the paper with the smart-pen, the pressure sensor of the smart-pen starts recording the pen movement on the micropatterned paper. The pen position on the paper is calculated from the acquired image through CMOS sensor of the e-pen. The position information of the pen is captured 85 times-per-second and transmitted to the evaluation and/or rehabilitation program operated on the PC through Bluetooth communication.

The position of the electronic pen is calculated based on the micropattern printed on the paper. A micropattern is a combination of several fine dots to express location information. One dot has a size of 50 to 80 μm , and 16 to 25 dots compose one location

information.

The position of the pen can be recognized with a resolution of 0.4 mm on paper, and the actual position of the pen is assumed to be the center point of the position pattern. By applying the mathematical interpolation algorithm, the maximum resolution is improved to 100 μ m (0.1 mm). Hence, in this study, the resolution of the pen was implemented as 0.1 mm.

Line tracing test

The cognitive ability evaluation method targeted in this paper is LTT, a type of the psychometric hepatic encephalopathy score proposed in 2001 (Randolph et al., 2009). It is used to evaluate the initial diagnosis of hepatic encephalopathy and the motor ability and accuracy of the arm. It is a prosecutor (Weissenborn et al., 2001). The test pattern for LTT consists of a 5-mm-thick straight and curved path, and the patient undergoing the test draws a line along the path to the destination without removing the pen. As the patient draws a line, the patient aims to avoid leaving or closing the path.

$$\text{score} = \tau \left(1 + \frac{\mu}{100}\right) \quad (\text{Equation 1})$$

The evaluation result is expressed as the line tracking score, (score), which is expressed as the product of the test time (τ) and the line tracking accuracy (μ) as shown in equation (1). The calculation method of line tracking accuracy (μ) is to divide the entire route into 365 areas and score points in each area to add up. The zones are divided into four cases and 0 points are awarded for not deviating from the route within the area. In addition, 1 points were given for closing the border, 2 points for leaving the border, and 3 points were given for disregarding the area.

LTT procedure

A paper with drawing of the LTT and patterns for real-time transmission of the drawing information to a computer with the application program was given to each participant and execution of the test was explained. The application program for cognitive ability evaluation receives location information from a smart-pen and displays it on the screen. The program evaluates the test results in real time based on the transmitted location information. Fig. 1 shows the execution of LTT using the developed smart-pen and the execution of the application program showing the result in real-time. In addition, the test results were scored so that a therapist or evaluator specialize in rehabilitation could immediately check the test results.

The right side shows what the subject marked, and the black color shows the line not marked by the subject. The e-system with program uniquely designed for LTT was implemented to enhance the rehabilitation effect by alerting the subject with vibration or LED light when the subject displays the line out of a certain range from the center of the line during the evaluation or ignores the line in order. The trajectory of the pen transmitted from the electronic pen to the PC through Bluetooth communication is automatically analyzed by the LTT-exclusive test-rehabilitation program, and the test time and LTT score are automatically analyzed and monitored by an evaluator. Same LTT test was conducted a week from the initial assessment for test-retest comparison.

Statistical analysis

Prior to the comparative assessments, the normality analysis was first performed using the Komogorov-Smirnov test for both data set. Both data sets were assessed to be normally distributed.

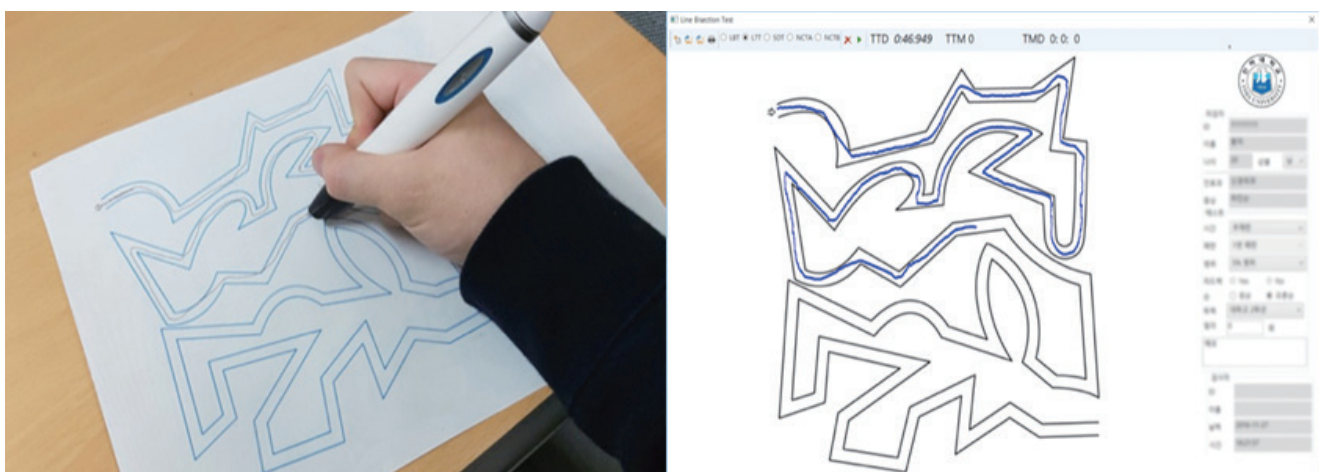


Fig. 1. Execution of the line tracing test with the novice e-pen system.

All statistical analysis was processed using MedCalc Statistical Software version 19.6 (Ostend, Belgium). The sample size of previous studies was first considered regarding the reliability of the LBT (Ferber and Karnath, 2001; Jee et al., 2015; Ku et al., 2009).

The interrater (system vs. golden-standard) and intrarater (test-retest) analyses were conducted on the assessed LTT test results evaluated by the system and the tester. The LTT test was conducted twice with a week of washout period between the repeated tests. The concordance correlation coefficients (CCCs) with confidence intervals (CIs) were used to measure the interrater reliability of the system. Intraclass correlations (ICCs) with CIs were used to assess the intrarater reliability of the repeated test results. As for the CCC and ICC coefficients, the results between 0.6 and 0.8 were considered 'substantial' and results greater than 0.8 were regarded as 'excellent' or 'near perfect' (Jee et al., 2015).

In addition, the effect sizes (Cohen *d*) between the comparison groups were also calculated. Effect size is a statistical method to compare the difference between two result means. When the effect size is 0.2–0.4, there is a small effect, when the effect size is about 0.5, the medium effect size (medium) is obtained. It shows that there is a large effect when it is a value (Rossetti et al., 2016). The CI range was 95%. Finally, the Bland-Altman plots between the system and tester first the first and second assessment trials were assessed. The Bland-Altman plots lineate mean differences between two assessment results with the degree of agreement between them (Jee et al., 2015). Closeness of the degree of agreements and CIs to zero indicate strong agreement between the two results. For all analyses, a significance level of $P \leq 0.05$ was set.

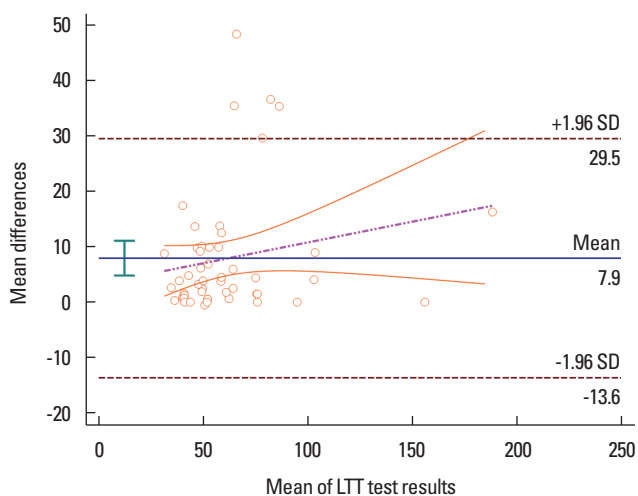


Fig. 2. Bland-Altman plot for comparison between the first e-system and first evaluator scores of the LTT test results. SD, standard deviation; LTT, line tracing test.

RESULTS

A total of 50 participants composed of 20 men and 30 women participated in this study. The participants were all right-handed with the mean age of 22.8 ± 2.57 years. The mean K-MMSE results for the first and second assessments were 29.48 and 29.38, respectively. The total assessment times for the first and second assessments were 56.48 and 62.59 seconds, respectively.

The CCC of the first assessment results between the system and the tester 1 with 95% CI was 0.89 (0.83–0.94) and second assessment result between the system and the tester 2 was 0.87 (0.79–0.92). In addition, CCC between the first system results and the second system results was 0.53 (0.37–0.66) and CCC between the first tester results and the second tester results was 0.64 (0.51–0.74).

The effect size between the first system and first tester assessed results was 0.27 and the effect size between second system and second tester assessed results was 0.27. Moreover, the effect size between the first system and second system results was 0.52 and the effect size between the first tester and second tester results was 0.40, respectively.

The Bland-Altman plots between the system and tester first the first and second assessment trials were assessed and shown in Figs. 2–5. The mean differences with CIs between the first system and tester assessed results were 7.9 (-13.6 and 29.5) (Fig. 2), between the second system and tester assessed results were 4.4 (-9.2 and 18.0) (Fig. 3), between the first system and second system assessed results were 12.6 (-29.8 and 55.0) (Fig. 4), and between the first

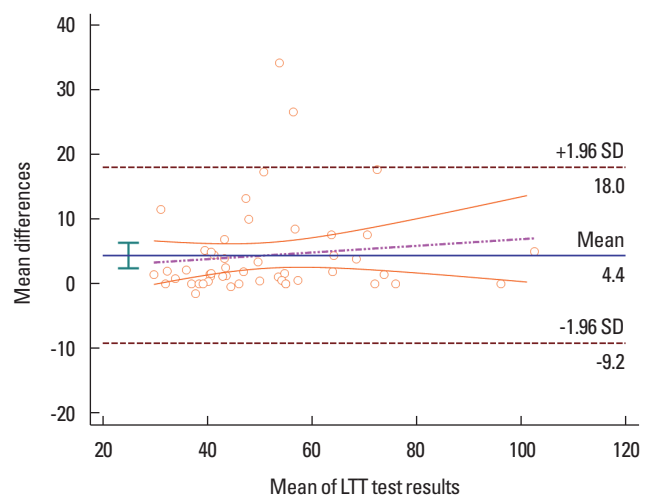


Fig. 3. Bland-Altman plot for comparison between the second e-system and second evaluator scores of the LTT test results. SD, standard deviation; LTT, line tracing test.

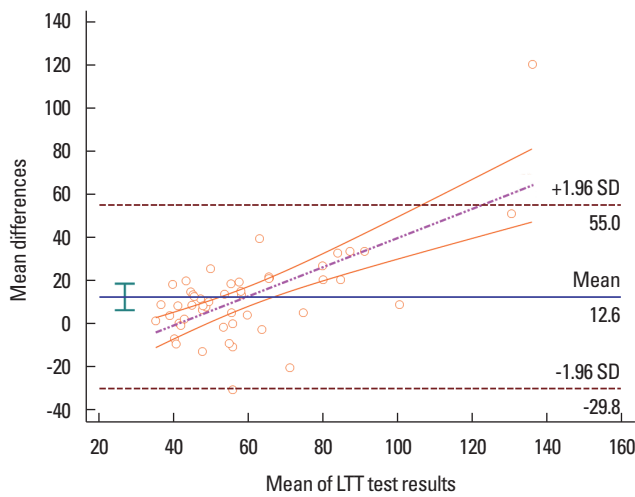


Fig. 4. Bland-Altman plot for comparison between the first e-system and second e-system scores of the LTT test results. SD, standard deviation; LTT, line tracing test.

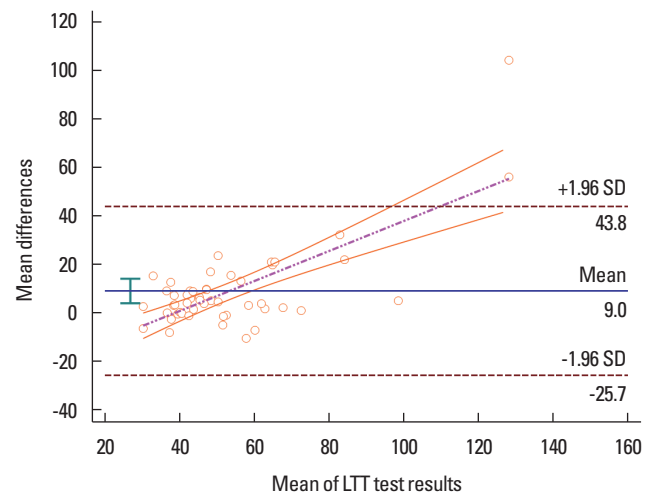


Fig. 5. Bland-Altman plot for comparison between the first evaluator and second evaluator scores of the LTT test results. SD, standard deviation; LTT, line tracing test.

Table 1. Validity assessment between the e-system and golden standard (evaluator) (n = 50)

Measured item	E-system	Evaluator	P-value	Effect sizes (Cohen d)	CCC (95% CI)
First test					
Accuracy rate (%)	96.5				
Evaluation duration (sec)	0	54.48 ± 54.26			
LTT scores	65.82 ± 31.53	58.00 ± 27.46	0.18	0.27	0.89 (0.83–0.94)
Retest					
Accuracy rate (%)	96.4				
Evaluation duration (sec)	0	62.59 ± 48.11			
LTT scores	61.22 ± 59.81	48.99 ± 15.37	0.17	0.27	0.87 (0.79–0.92)

Values are presented as mean ± standard deviation.

CCC, concordance correlation coefficient; CI, confidence interval; LTT, line tracing test; E-system, novice electronic pen-based LTT evaluation system; accuracy rate (%), LTT score/maximum LTT score × 100; evaluation duration (sec), time that took the evaluator to complete assessment.

tester and second tester assessed results were 9.0 (-25.7 and 43.8) (Fig. 5).

DISCUSSION

This study examined the feasibility of a smart-type LTT system (e-system) made for evaluating cognitive dysfunction, rehabilitation treatment, and cognitive function evaluation. The system that automatically evaluates LTT, a type of psychometric test tools used to evaluate cognitive dysfunction in patients, was developed and utilized for this study. Following the assessment, the LTT results were evaluated automatically by the novice e-system and manually by an evaluator.

In order to verify the accuracy and effectiveness of the developed system, the LTT test was conducted on 20 men and 30 women

participated. The novice evaluation tool using an electronic pen in this paper was observed to be within the error range of the existing manual evaluation method. Moreover, the novice system ensured the measurement reliability as an assessment tool. Repeated tests were conducted for the e-system and evaluator comparisons. The average time the evaluator spent evaluating the LTT results were 54.7 ± 54.3 and 62.6 ± 48.1 sec for the first and second tests. This indicates that the manual evaluation time of about 1 min. In the case of this study, the novice evaluation system showed little-to-no evaluation time as shown in Table 1. Such time results were similar to previous study that performed with similar type of psychometric test (Jee et al., 2015).

In order to observe the feasibility, the test results of the novice system were compared with the results assessed by the golden standard. As shown in Table 1, mean comparisons were not sig-

nificant. In addition, the first and second LTT tests showed small effect sizes of 0.27 for both trials. As indicated by previous study, effect size between 0.2 and 0.4 were considered as small effect size, effect size of 0.5 was considered as medium effect size, and effect size greater than 0.5 was considered as large effect size (Rossetti et al., 2016). In addition to effect sizes, CCCs were also calculated. E-system and evaluator comparisons for the first and second LTT test results were observed. As shown in Table 1, CCCs for first and second comparisons were close to 0.9. The correlation coefficients of previous studies showed that results between 0.6 and 0.8 as significantly correlated and 0.8 or above as very highly correlated.

Finally, Bland-Altman plots showed strong degree of agreement for the assessment and re-assessment comparisons between the e-system and the evaluator. The limit of agreement provides information on the degree of error and repeatability (Abu-Arafeh et al., 2016; Bland and Altman, 1999). Smaller the scatters of the differences along the central line indicate high repeatability and greater degree of agreement. The results of this study show the smallest degree of dispersion of the scattered mean differences for the second LTT assessment results. Moreover, the Bland-Altman plots between the e-system and evaluator were comparatively closer to each other than the comparisons between the first and second test and retest results even with the same assessment methods. Such observations indicate that the comparison of two repeated measures leads to differences in true results despite the identical subjects and environment (Bland and Altman, 2007; Pagnacco et al., 2015). As shown in Table 1, the results between the first and second tests showed significant differences between the results despite the same assessment methods applied. In addition to all the closeness comparisons between the e-system and evaluator, the accuracy rates for both tests were 97% on the average. All the comparative results between the e-system and the golden-standard method conducted by an evaluator showed high feasibility of the novice system.

The electronic pen-based system for evaluating the LTT using a paper-and-pen can be widely used in hospitals without using different interfaces such as computer screen. However, due to the time-consuming assessment and evaluation time, computer-aided assessment tools are being suggested (Jee et al., 2015; Luo et al., 2019; Mardini et al., 2008). According to a review paper, the diagnostic and screening time of different tests were between 10 to 30 min (Luo et al., 2019). This novice system will provide the advantage of allowing professional medical professionals to use more time for patients by remarkably shortening evaluation and storage

time along with reliable results.

There were some limitations to the study. First, assessment by the e-system showed errors of not counting excessively outlying lines that could not be identified by the e-pen due to the excessively tilted angle of the pen during some drawing movements. Such limitation could lead to greater error with the cognitively impaired patients due to limitation of limb control. Another limitation would be the subjects of this study. All the test subjects were physically and mentally healthy young subjects. The drawing patterns would be significantly different from the cognitively impaired patients. Greater variations would be expected from the patient. However, testing the feasibility of the novice e-system with the healthy subjects could obtain the opportunity to check the feasibility prior to proceeding with fragile patients.

As future studies, we plan to conduct clinical studies targeting actual patients, and through this, it is expected that the utility of the system can be empirically presented with clinical application.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

- Abu-Arafeh A, Jordan H, Drummond G. Reporting of method comparison studies: a review of advice, an assessment of current practice, and specific suggestions for future reports. *Br J Anaesth* 2016;117:569-575.
- Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat* 2007;17:571-582.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135-160.
- de Joode E, van Heugten C, Verhey F, van Boxtel M. Efficacy and usability of assistive technology for patients with cognitive deficits: a systematic review. *Clin Rehabil* 2010;24:701-714.
- Ferber S, Karnath HO. How to assess spatial neglect—line bisection or cancellation tasks? *J Clin Exp Neuropsychol* 2001;23:599-607.
- Jee H, Kim J, Kim C, Kim T, Park J. Feasibility of a semi-computerized line bisection test for unilateral visual neglect assessment. *Appl Clin Inform* 2015;6:400-417.
- Jeong JY, Jun DW, Bai D, Kim JY, Sohn JH, Ahn SB, Kim SG, Kim TY, Kim HS, Jeong SW, Cho YK, Song DS, Kim HY, Jung YK, Yoon EL. Validation of a paper and pencil test battery for the diagnosis of minimal hepatic encephalopathy in Korea. *J Korean Med Sci* 2017;32:1484-1490.

- Kang IW, Beom IG, Cho JY, Son HR. Accuracy of Korean-mini-mental status examination based on Seoul neuro-psychological screening battery II results. *Korean J Fam Med* 2016;37:177.
- Kerai JH, Bracewell RM, Hindle JV, Leek EC. Visuospatial transformation impairments in Parkinson's disease. *J Clin Exp Neuropsychol* 2012;34:1053-1064.
- Ku J, Lee JH, Han K, Kim SI, Kang YJ, Park ES. Validity and reliability of cognitive assessment using virtual environment technology in patients with stroke. *Am J Phys Med Rehabil* 2009;88:702-710.
- Li SW, Wang K, Yu YQ, Wang HB, Li YH, Xu JM. Psychometric hepatic encephalopathy score for diagnosis of minimal hepatic encephalopathy in China. *World J Gastroenterol* 2013;19:8745-8751.
- Luo M, Ma P, Li L, Cao WK. Advances in psychometric tests for screening minimal hepatic encephalopathy: from paper-and-pencil to computer-aided assessment. *Turk J Gastroenterol* 2019;30:398-407.
- Mardini H, Saxby BK, Record CO. Computerized psychometric testing in minimal encephalopathy and modulation by nitrogen challenge and liver transplant. *Gastroenterology* 2008;135:1582-1590.
- Pagnacco G, Carrick FR, Wright CH, Oggero E. Between-subjects differences of within-subject variability in repeated balance measures: consequences on the minimum detectable change. *Gait Posture* 2015;41:136-140.
- Randolph C, Hilsabeck R, Kato A, Kharbanda P, Li YY, Mapelli D, Ravdin LD, Romero-Gomez M, Stracciari A, Weissenborn K; International Society for Hepatic Encephalopathy and Nitrogen Metabolism (ISHEN). Neuropsychological assessment of hepatic encephalopathy: ISHEN practice guidelines. *Liver Int* 2009;29: 629-635.
- Rossetti MA, Piryatinsky I, Ahmed FS, Klinge PM, Relkin NR, Salloway S, Ravdin LD, Brenner E, Malloy PF, Levin BE, Broggi M, Gavett R, Maniscalco JS, Katzen H. Two novel psychomotor tasks in idiopathic normal pressure hydrocephalus. *J Int Neuropsychol Soc* 2016;22:341-349.
- Weissenborn K, Ennen JC, Schomerus H, Rückert N, Hecker H. Neuropsychological characterization of hepatic encephalopathy. *J Hepatol* 2001;34:768-773.