

# Test–retest reliability and convergent validity of the Test of Nonverbal Intelligence-Fourth Edition in patients with schizophrenia.

**Kuan-Wei Chen**

Kaohsiung Municipal Kai-Syuan Psychiatric Hospital

**Ya-Chen Lee**

Asia University

**Tzu-Ying Yu**

I-Shou University

**Li-Jung Cheng**

Kaohsiung Municipal Kai-Syuan Psychiatric Hospital

**Chien-Yu Chao**

Kaohsiung Municipal Kai-Syuan Psychiatric Hospital

**Ching-Lin Hsieh** (✉ [clhsieh@ntu.edu.tw](mailto:clhsieh@ntu.edu.tw))

National Taiwan University <https://orcid.org/0000-0001-9468-9969>

---

## Research article

**Keywords:** cognition, fluid intelligence, intelligence, psychometrics, schizophrenia

**Posted Date:** November 4th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-27553/v3>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on January 13th, 2021. See the published version at <https://doi.org/10.1186/s12888-021-03041-4>.

# Abstract

**Background:** Fluid intelligence deficits affect executive functioning and social behaviors in patients with schizophrenia. To help clinicians manage fluid intelligence deficits, a psychometrically sound measure is needed. The purposes of this study were to examine the test–retest reliability and convergent validity of the Test of Nonverbal Intelligence-Fourth Edition (TONI-4) assessing fluid intelligence in patients with schizophrenia.

**Methods:** A total of 103 patients with stable condition were assessed with the TONI-4 twice with a 4-week interval to examine the test–retest reliability. We further used the Montreal Cognitive Assessment (MoCA) and the Tablet-Based Symbol Digit Modalities Test (T-SDMT) to examine the convergent validity of the TONI-4.

**Results:** The intra-class correlation coefficient was 0.73 for the TONI-4. The percentages of standard error of measurement and minimal detectable change for the TONI-4 were 5.1% and 14.2%, respectively. The practice effect of the TONI-4 was small (Cohen's  $d = -0.03$ ). Convergent validity showed small to moderate significant correlations between the TONI-4 and the MoCA as well as the T-SDMT ( $r = 0.35$ ,  $p = .011$  with the T-SDMT and  $r = 0.61$ ,  $p < .001$  with the MoCA). The results demonstrated that the TONI-4 had good test–retest reliability, limited random measurement error, and a trivial practice effect. The convergent validity of the TONI-4 was good.

**Conclusions:** These findings indicate that the TONI-4 has potential to be a reliable and valid assessment of fluid intelligence in patients with schizophrenia.

## Background

Fluid intelligence can be defined as the ability to think logically and solve problems in novel situations [1, 2]. Conceptually, fluid intelligence has been linked to executive functioning and complex social behavior [3, 4]. Fluid intelligence is a critical cognitive ability affecting a wide variety of daily activities [5, 6]. Fluid intelligence deficits are common in patients with schizophrenia, and these deficits are often associated with cognitive impairment in this group [7-12]. The deficits of fluid intelligence in patients with schizophrenia are associated with difficulties in daily independent functioning [10, 11, 13]. Moreover, low fluid intelligence, which is included in the premorbid intelligence exhibited by patients with schizophrenia, precedes the first psychotic episode and appears to be related to the risk for schizophrenia [4, 8, 10, 14-17]. To help clinicians manage patients' fluid intelligence, clinicians and researchers have to administer reliable and valid assessments of such deficits.

Three assessments are commonly used to assess fluid intelligence [4, 18-20]. They are the Comprehensive Test of Nonverbal Intelligence–Second Edition (CTONI-2) [21, 22], the Raven Advanced Progressive Matrices Test (RAPM) [23, 24], and the Test of Nonverbal Intelligence–Fourth Edition (TONI-4) [25]. However, some items of the CTONI-2 have cultural bias. For example, one item includes pictures related to American football or faces of Caucasians. In contrast, both the RAPM and the TONI-4 are

administered with geometric patterns, which are not culturally dependent. Comparing the administration time, the CTONI-2 and the RAMP usually take an average of 40–60 minutes to finish, while the TONI-4 can be finished within 15 minutes. For time-pressed clinicians, the TONI-4 has great potential for assessing fluid intelligence in patients with schizophrenia.

Some supportive evidence on the psychometric properties has been found for the TONI-4 (e.g., sufficient test–retest reliability and construct validity) in healthy groups [25, 26]. However, the TONI-4 has not yet been validated in patients with schizophrenia. Because psychometric properties are generally sample dependent [27, 28], psychometric studies are needed to confirm whether the TONI-4 is reliable and valid in patients with schizophrenia. Particularly, sufficient psychometric properties (e.g., test–retest reliability, practice effect, random measurement error, and validity) are required for a measure to ensure its clinical utility for repeated assessments in patients with schizophrenia.

The current study aimed to examine the test–retest reliability, practice effect, random measurement error, and convergent validity of the TONI-4 in patients with schizophrenia. The results of the study should help clinicians and researchers determine the utility of the TONI-4 when applied to patients with schizophrenia.

## Methods

### Participants

We recruited participants via convenience sampling from a psychiatric hospital in southern Taiwan between June 2017 and April 2018. Patients were included in this study if they met the following criteria: (1) diagnosis of schizophrenia according to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition [29]. DSM-5 criteria for schizophrenia was assessed and validated by board-certified psychiatrists and supported by clinical observations and interviews during hospitalization, past medical records, and information provided by main caregivers, (2) age  $\geq$  20 years, and (3) stable use and dosage of antipsychotic medication for at least one month prior to recruitment. The exclusion criteria were (1) diagnosis of other neurological or psychiatric diseases affecting cognition (e.g., stroke or depression), (2) another severe medical condition or psychiatric disorder that required treatment during the study, or (3) unstable severity of symptoms [specifically, a change in score of more than 2 on the Clinical Global Impressions Scale–Severity (CGI-S)] [30].

This study was approved by the Institutional Review Board of the local hospital. All participants signed consent forms before participating in this study.

### Procedure

This study was comprised of three assessments with 2-week intervals between adjacent assessments (i.e., early, middle and late assessments). At the early and late assessments, the participants completed alternate forms of the TONI-4 (i.e., Form A at the early assessment and Form B at the late assessment) at a four-week interval. At the middle assessment, we administered the Tablet-Based Symbol Digit

Modalities Test (T-SDMT) [31] and the Montreal Cognitive Assessment (MoCA) [32]. All assessments were administered by a trained occupational therapist, using standardized protocols, forms, and manuals. In addition, the CGI-S was administered in each session to confirm that the participants' symptoms did not change during the study period. We collected the patients' demographic characteristics from chart review.

## Measures

### *TONI-4*

The TONI-4 is designed to assess fluid intelligence in individuals aged 6 years to 89 years and 11 months. The TONI-4 has alternate forms (i.e., Form A and Form B) to reduce the practice effect [25]. Items for Form A can be found on one side of the picture book, and items for Form B are found on the reverse side. The two forms are not interchangeable. Each item is composed of a sequence of abstract figures with a figure missing from the sequence. Each sequence includes one or more attributes, such as shape, position, direction, rotation, contiguity, shading, size, and movement. Items ascend in level of difficulty as more attributes are added. When three of the five consecutive items are incorrectly answered, the test is terminated. The items are scored dichotomously: Correct answers earn one point and incorrect answers earn zero points. The rater writes the score on the answer sheet. The TONI-4 is norm referenced and yields an index, which is a standardized score (quotient) with a mean of 100 and a standard deviation of 15. Higher index scores indicate better fluid intelligence [25].

### *T-SDMT*

The T-SDMT was developed from the SDMT to assess processing speed [31]. This test includes 9 different symbols, each associated with a number (1–9), presented to the examinee on a tablet computer screen (i.e., an iPad). All trials are conducted with the tablet in landscape orientation, held in place by a case that is adjusted to a 30-degree tilting angle. To respond to each item, the participant is required first to look at the symbol in the center of the screen, then to search for the corresponding number in the table at the top of the screen, and finally to choose the corresponding number on a 3-by-3 grid at the bottom of the screen. The tablet computer automatically records the number of correct answers during the test. A higher number of correct answers indicates better performance of processing speed. The T-SDMT has acceptable psychometric properties in patients with schizophrenia [31].

### *MoCA*

The MoCA briefly measures overall cognitive functioning, including orientation, memory, visuospatial skills, executive functioning, language, and attention [32]. The total scores range between 0 and 30, and higher scores indicate better cognitive functioning. The total score (including the addition of one point for examinees with 12 or fewer years of education) is used for analysis. The MoCA has demonstrated high sensitivity as a cognitive screening test for severe mental illness [33].

### *CGI-S*

The CGI-S assesses symptom severity on a 7-point scale (1–7) [30]. One point on the CGI-S represents that a patient is not ill, and 7 points represents most severely ill. We used the CGI-S to examine whether the symptom severity of the participants was stable during the study period.

## **Statistical Analyses**

### *Test–retest reliability*

Test–retest reliability was estimated using the intra-class correlation coefficient (ICC) between the early and late assessments, on the basis of a two-way random-effects model with absolute agreement [34]. The following criteria were used to interpret ICC values: an ICC value  $\geq 0.80$  indicated excellent test–retest reliability; 0.60–0.79, good; 0.40–0.59, moderate; and  $< 0.40$ , poor [35].

The standard error of measurement (SEM) is an index of random measurement error that can be used to present the precision of individual scores [36]. The SEM% was calculated by dividing the SEM by the mean of the early assessment score and then multiplying the result by 100% (SEM%). An SEM% of less than 10% is considered to indicate limited random measurement error for a measure [37].

We also calculated the minimal detectable change (MDC) and MDC percentage (MDC%) to examine the change between adjacent assessments that could be considered as a real change (beyond the score change caused by random measurement error) at the 95% confidence level. The MDC% was calculated by dividing the MDC by the mean of the early assessment score and then multiplying the result by 100% [38].

In addition, the agreement between test–retest measurements was analyzed by Bland–Altman plots with 95% limits of agreement (LOA) [39]. In these plots, the differences ( $d$ ) between each pair of assessments were presented against the average value for each pair of assessments. To examine whether heteroscedasticity existed, Pearson’s correlation coefficient ( $r$ ) was used to calculate the correlation between the absolute value of the difference of two assessments and the mean score of two assessments [40]. When Pearson’s  $r$  was  $\geq 0.3$  or  $\leq -0.3$ , it meant that the absolute value of the difference was related to the mean score of two assessments, and that there was heteroscedasticity [41]. In other words, the higher the assessment score, the greater ( $r \geq 0.3$ ) or smaller ( $r \leq -0.3$ ) the difference between the two assessments.

Effect size (Cohen’s  $d$ ) was used to estimate the magnitudes of practice effects due to repeated assessments of the TONI-4. An effect size  $\geq 0.80$  was considered as a large practice effect; 0.50–0.79, medium; 0.20–0.49, small; and  $< 0.20$ , trivial [42].

To further examine whether the findings were consistent across participants’ genders and ages, subgroup analysis was performed. We stratified the participants by gender and three age bands (i.e., 20–39, 40–49, and 50–70) individually.

### *Convergent validity*

Convergent validity was examined by correlating the scores of the TONI-4 at the early assessment with those of the MoCA and the T-SDMT using Pearson's  $r$ . We hypothesized that we would find moderate correlations between the scores of the TONI-4 and the MoCA (i.e., fluid intelligence and cognition) [43], and that small correlations would be found between the scores of the TONI-4 and the T-SDMT (i.e., fluid intelligence and processing speed) [3, 18].

## Results

We recruited 106 patients with schizophrenia who were eligible for the study. Of these, 103 participants completed all assessments. About half of the participants were male (50.5%), and the mean age was 46.7 years. The demographic and clinical characteristics of the participants are shown in Table 1. The early and late assessments scores of the TONI-4, on average, were very similar (92.4 and 91.9), indicating that the participants had slight impairment of fluid intelligence. In addition, the mean score of the MoCA was 23.3, indicating that our participants, on average, had mild cognitive impairment. The mean score of the T-SDMT was 32.1, indicating that the processing speed of most participants was impaired.

### Test–retest reliability

Table 2 shows the results of the test–retest reliability analyses. The ICC of the TONI-4 was 0.73 (95% confidence interval: 0.62 to 0.81).

The SEM (SEM%) and MDC (MDC%) of the TONI-4 scale were 4.7 (5.1%) and 13.1 (14.2%) points, respectively. The results were smaller than our preset criterion.

In Figure 1, the LOAs ranged from -14.6 to 13.6 points. Pearson's  $r$  between the absolute value of the difference of the early and late assessments and the mean score of the early and late assessments was 0.31.

Analysis of the practice effect revealed that the effect size of score change in the TONI-4 was small (Cohen's  $d = -0.03$ ) between the early and late assessments.

To further examine whether the aforementioned findings were consistent across participants' genders (male and females) and ages (20–39, 40–49, and 50–70), sub-group analysis was performed. The results showed that the ICCs (0.64–0.82), SEM%s (4.2%–6.3%), MDC%s (11.7%–17.3%), and Cohen's  $d$ s (-0.29–0.13) of the TONI-4 were similar across all sub-groups.

### Convergent validity

The index scores of the TONI-4 were moderately correlated with the scores of the MoCA ( $r = 0.61$ ,  $p < .001$ ,  $n = 96$ ), whereas a small correlation was found between the scores of the TONI-4 and the T-SDMT ( $r = 0.35$ ,  $p = .011$ ,  $n = 51$ ). In addition, there were no significant differences in the scores of the TONI-4 between the participants who had and those who had not been assessed with the MOCA ( $t = -4.82$ ,  $p = 0.631$ ) and the T-SDMT ( $t = -4.29$ ,  $p = 0.669$ ).

# Discussion

## Test–retest reliability

A measure with sufficient test–retest reliability ensures that users can obtain reproducible scores. Good test–retest reliability was found for the repeated assessments of the TONI-4. Moreover, the test–retest reliabilities were similar across the gender and age sub-groups. Accordingly, the TONI-4 has generally good test–retest reliability, which may not be affected by examinees' gender and age, and it can be used in repeated assessments. In comparison with previous studies, the test–retest reliability of our study was slightly lower than those found for healthy controls ( $r = 0.82–0.93$ ) [25] and was consistent with those of other cognitive assessments examining patients with schizophrenia [44, 45]. There are three possible reasons for the slightly lower ICC of the TONI-4. First, the test–retest reliability was estimated by Pearson correlation coefficients in the previous studies, which tends to overestimate reliability [34]. Second, alternate forms (i.e., Forms A and B) were used in this study, which may have resulted in more variation compared to using the same form as previous studies [25]. Third, the heterogeneity of our sample appeared limited. In particular, the variances of the TONI-4 in this study (SDs = 9.1 and 9.9) were smaller than those of a previous study (SDs = 13–15) [25], which may have underestimated the ICC values in this study [46]. In summary, our findings indicate that the TONI-4 appears to be reliable for repeatedly assessing fluid intelligence in patients with schizophrenia.

We found that the SEM% was far below our preset criterion. Furthermore, the SEM%s were generally consistent across the gender and age sub-groups. These findings suggest that the TONI-4 has limited random measurement error. Our findings are consistent with those in previous studies examining healthy groups, where the SEM% were 4.0–5.5% [25]. These findings support that the random measurement error is similar in patients with schizophrenia and in healthy adults. Therefore, the scores of the TONI-4 tend to be stable in patients with schizophrenia.

In addition, MDC can be viewed as the threshold for a statistically significant change for individual patients in clinical and research settings [47]. Conceptually, a change exceeding the MDC of the first assessment can be interpreted as a real improvement with the corresponding certainty (e.g., 95%). Thus, a fixed MDC value can be used to interpret the change scores for patients with different levels of fluid intelligence. However, we found that the association between the absolute value of the difference of the early and late assessments and the mean score of the early and late assessments (Pearson's  $r = 0.31$ ) was above 0.30, implying the existence of heteroscedasticity [41]. That is, the absolute difference and the mean of the early and late assessments increased simultaneously. Accordingly, a fixed value of MDC is not appropriate for different levels of fluid intelligence.

In such assessments with heteroscedasticity, the MDC% is more suitable than the MDC for interpreting a true change for a patient [48]. That is, as seen in this study, the MDC value can be adjusted based on the MDC% and the patient's early assessment score. Specifically, the MDC% (14.2%) of the TONI-4 can be multiplied by the patient's early assessment score to achieve an adjusted MDC value. For example, a patient with a score of 92 points at the early assessment requires an improvement of more than 13.1

points ( $92 \times 0.142$ ) to indicate a true change. These adjustments can help clinicians and researchers interpret the score changes on the TONI-4 of an individual patient after intervention and then develop further treatment plans accordingly.

We found that the scores between the early and late assessments had almost no change. In addition, those values were similar across the sub-groups of examinees' gender and age. These findings indicate that the scores of the TONI-4 do not systematically increase given that the early assessment (or practice) has already been completed. Our findings are consistent with those in a previous study, where the change scores within one-to-two-week intervals were small (effect size = 0.00–0.07) [25]. The trivial practice effect may have been due to the use of alternate forms (i.e., Forms A and B) [49, 50]. However, using alternate forms may lead to underestimation of the practice effect as compared to using a single form. In this study, all participants were administered the forms in a fixed order (i.e., Form A first and Form B second). The fixed order design was used because previous findings had indicated that test–retest reliability is not affected by the order effect [26]. Thus, clinicians could use alternate forms of the TONI-4 in their routine repeated assessments to effectively minimize practice effects.

### **Convergent validity**

We found that the scores of the TONI-4 were moderately correlated with those of the MoCA and significantly correlated with those of the T-SDMT, supporting our hypotheses. Thus, good convergent validity was demonstrated for the TONI-4. Our results support the validity of the TONI-4 for assessing fluid intelligence in patients with schizophrenia.

This study had two merits. First, the sample size (103 participants) was relatively large. A large sample size tends to provide robust estimates, which improves the generalizability of our findings [51]. Second, we used alternate forms of the TONI-4. Due to this study design, the practice effects of the TONI-4 were well controlled, so its utility in repeated assessments was confirmed.

### **Implications**

The fluid intelligence encompasses the ability to think logically and solve problems in novel situations, which is a critical cognitive ability affecting clients' performance on a wide variety of daily activities. Knowledge and evidence of the test–retest reliability and convergent validity of the TONI-4 help clinicians select a measure for assessing fluid intelligence in patients with schizophrenia.

- The TONI-4 appears reliable for repeatedly assessing the fluid intelligence in patients with schizophrenia.
- Due to heteroscedasticity of the TONI-4, an adjusted MDC, the patient's early assessment score multiplied by the MDC% (14.2%), is suggested for use in determining whether the change in score of a patient is outside the range of random measure error.
- The good convergent validity of the TONI-4 provides a preliminary basis to support its utility for assessing fluid intelligence in patients with schizophrenia.

## Study limitations

Two limitations of this study should be noted. First, the study sample was a convenience sample recruited from a psychiatric center in southern Taiwan. In addition, our participants, on average, had slightly impaired fluid intelligence (the mean score of the TONI-4 was 92.4 points at the early assessment). The above sampling limitations might have affected the generalizability of our findings. Second, we used alternate forms to examine the test–retest reliability of the TONI-4. Thus, our results on test–retest reliability might not be generalizable to single-form assessment of the TONI-4. Using alternate forms may lead to underestimation of the test–retest reliability as compared to using a single form.

## Conclusions

We found good test–retest reliability and good convergent validity of the TONI-4 in patients with schizophrenia. These findings provide preliminary evidence supporting the utility of the TONI-4 in patients with schizophrenia.

## Abbreviations

CTONI-2: Comprehensive Test of Nonverbal Intelligence–Second Edition; RAPM: Raven Advanced Progressive Matrices Test; TONI-4: Test of Nonverbal Intelligence–Fourth Edition; DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition; CGI-S: Clinical Global Impressions Scale–Severity; T-SDMT: Tablet-Based Symbol Digit Modalities Test; MoCA: Montreal Cognitive Assessment; ICC: Intra-class correlation coefficient; SEM: Standard error of measurement; MDC: Minimal detectable change; LOA: Limits of agreement; SD: standard deviation

## Declarations

### Ethics approval and consent to participate

Ethical approval for this study was granted by the Institutional Review Board (IRB) of the Kaohsiung Municipal Kai-Syuan Psychiatric Hospital. All subjects were informed about the study and signed consent forms before participating in this study.

### Consent for publication

Not Applicable

### Availability of data and materials

The (anonymized) datasets analyzed during the current study are available from the first author on reasonable request.

### Competing interests

The authors declare that they have no competing interests.

## Funding

This study was supported by the Kaohsiung Municipal Kai-Syuan Psychiatric Hospital (grant number KSPH-2016-03). The funder had no role in study design and collection, analysis, and interpretation of data, decision to publish and in writing the manuscript.

## Author contributions

KWC and CLH designed the study; LJC and CYC collected the data; KWC and CLH analyzed the data and wrote the first draft; KWC, YCL and TYY revised the manuscript; all authors gave insightful comments and improved the manuscript. All authors approved the final version of the manuscript.

## Acknowledgements

The authors would like to thank all the participants in the study.

## References

1. Cattell RB. Theory of fluid and crystallized intelligence: A critical experiment. *J Educ Psychol.* 1963;54(1):1-22.
2. Horn JL, Cattell RB. Refinement and test of the theory of fluid and crystallized general intelligences. *J Educ Psychol.* 1966;57(5):253-270.
3. Roca M, Manes F, Cetkovich M, Bruno D, Ibanez A, Torralva T, Duncan J. The relationship between executive functions and fluid intelligence in schizophrenia. *Front Behav Neurosci.* 2014;8:46.
4. Huepe D, Roca M, Salas N, Canales-Johnson A, Rivera-Rei AA, Zamorano L, Concepcion A, Manes F, Ibanez A. Fluid intelligence and psychosocial outcome: from logical problem solving to social adaptation. *PLoS One.* 2011;6(9):e24858.
5. Gray JR, Thompson PM. Neurobiology of intelligence: science and ethics. *Nat Rev Neurosci.* 2004;5(6):471-482.
6. Jaeggi SM, Buschkuhl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. *Proc Natl Acad Sci U S A.* 2008;105(19):6829-6833.
7. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology.* 1998;12(3):426-445.
8. Woodberry KA, Giuliano AJ, Seidman LJ. Premorbid IQ in schizophrenia: a meta-analytic review. *Am J Psychiatry.* 2008;165(5):579-587.
9. Chandler D, Dragovic M, Cooper M, Badcock JC, Mullin BH, Faulkner D, Wilson SG, Hallmayer J, Howell S, Rock D *et al.* Impact of Neuritin 1 (NRN1) polymorphisms on fluid intelligence in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet.* 2010;153B(2):428-437.

10. Kievit RA, Davis SW, Griffiths J, Correia MM, Cam C, Henson RN. A watershed model of individual differences in fluid intelligence. *Neuropsychologia*. 2016;91:186-198.
11. Sternberg RJ, Wagner RK. Practical intelligence: Nature and origins of competence in the everyday world. New York, NY: CUP Archive; 1986.
12. Van Rheenen TE, Cropley V, Fagerlund B, Wannan C, Bruggemann J, Lenroot RK, Sundram S, Weickert CS, Weickert TW, Zalesky A *et al*. Cognitive reserve attenuates age-related cognitive decline in the context of putatively accelerated brain ageing in schizophrenia-spectrum disorders. *Psychol Med*. 2020;50(9):1475-1489.
13. Tucker-Drob EM. Global and domain-specific changes in cognition throughout adulthood. *Dev Psychol*. 2011;47(2):331-343.
14. Snitz BE, Macdonald AW, 3rd, Carter CS. Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: a meta-analytic review of putative endophenotypes. *Schizophr Bull*. 2006;32(1):179-194.
15. Blair C. How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behav Brain Sci*. 2006;29(2):109-125.
16. Caspi A, Reichenberg A, Weiser M, Rabinowitz J, Kaplan Z, Knobler H, Davidson-Sagi N, Davidson M. Cognitive performance in schizophrenia patients assessed before and following the first psychotic episode. *Schizophr Res*. 2003;65(2-3):87-94.
17. Khandaker GM, Barnett JH, White IR, Jones PB. A quantitative meta-analysis of population-based studies of premorbid intelligence and schizophrenia. *Schizophr Res*. 2011;132(2-3):220-227.
18. Shelton JT, Elliott EM, Matthews RA, Hill BD, Gouvier WD. The relationships of working memory, secondary memory, and general fluid intelligence: working memory is special. *J Exp Psychol Learn Mem Cogn*. 2010;36(3):813-820.
19. McGill RJ. Investigation of the factor structure of the Comprehensive Test of Nonverbal Intelligence—Second Edition (CTONI-2) using exploratory factor analysis. *J Psychoeduc Assess*. 2016;34(4):339-350.
20. Rossen EA, Shearer DK, Penfield RD, Kranzler JH. Validity of the Comprehensive Test of Nonverbal Intelligence (CTONI). *J Psychoeduc Assess*. 2005;23(2):161-172.
21. Hammill DD, Pearson N. Comprehensive test of nonverbal intelligence. In: *Handbook of Nonverbal Assessment*. edn. Boston, MA: Springer; 2017: 167-184.
22. Hammill DD, Pearson NA, Wiederholt JL. Comprehensive test of nonverbal intelligence (CTONI). Austin, TX: Pro-ed 1997.
23. Raven JC. Manual for Raven's progressive matrices and vocabulary scales. London: HK Lewis; 1983.
24. Raven JC. Guide to the standard progressive matrices: Sets A, B, C, D and E. London: Lewis & Co; 1960.
25. Brown L, Sherbenou RJ, Johnsen SK. Test of nonverbal intelligence: TONI-4. Austin, TX: Pro-ed; 2010.

26. Ritter N, Kilinc E, Navruz B, Bae Y. Test Review: Test of Nonverbal Intelligence-4 (TONI-4). *J Psychoeduc Assess*. 2011;29(5):484-488.
27. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):iii, ix-x, 1-177.
28. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York, NY: McGraw-Hil; 1994.
29. American Psychiatric Association. *Desk reference to the diagnostic criteria from DSM-5®*. Arlington, VA: American Psychiatric Publishing, Inc.; 2014.
30. Haro J, Kamath S, Ochoa S, Novick D, Rele K, Fargas A, Rodriguez M, Rele R, Orta J, Kharbeng A. The Clinical Global Impression–Schizophrenia scale: A simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatr Scand*. 2003;107(416):16-23.
31. Tung LC, Yu WH, Lin GH, Yu TY, Wu CT, Tsai CY, Chou W, Chen MH, Hsieh CL. Development of a Tablet-based symbol digit modalities test for reliably assessing information processing speed in patients with stroke. *Disabil Rehabil*. 2016;38(19):1952-1960.
32. Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.
33. Musso MW, Cohen AS, Auster TL, McGovern JE. Investigation of the Montreal Cognitive Assessment (MoCA) as a cognitive screener in severe mental illness. *Psychiatry Res*. 2014;220(1-2):664-668.
34. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep*. 1966;19(1):3-11.
35. Bushnell CD, Johnston DC, Goldstein LB. Retrospective assessment of initial stroke severity: comparison of the NIH Stroke Scale and the Canadian Neurological Scale. *Stroke*. 2001;32(3):656-660.
36. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19(1):231-240.
37. Flansbjerg UB, Holmback AM, Downham D, Lexell J. What change in isokinetic knee muscle strength can be detected in men and women with hemiparesis after stroke? *Clin Rehabil*. 2005;19(5):514-522.
38. Huang SL, Hsieh CL, Wu RM, Tai CH, Lin CH, Lu WS. Minimal detectable change of the timed" up & go" test and the dynamic gait index in people with Parkinson disease. *Phys Ther*. 2011;91(1):114-121.
39. Bland JM, Altman DG. Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *Lancet*. 1986;1(8476):307-310.
40. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135-160.
41. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*. 1998;26(4):217-238.

42. Cohen J. Statistical power analysis for the behavioral sciences, 2 edn. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
43. Cochrane M, Petch I, Pickering AD. Aspects of cognitive functioning in schizotypy and schizophrenia: evidence for a continuum model. *Psychiatry Res.* 2012;196(2-3):230-234.
44. Hahn E, Vollath A, Ta TT, Hahn C, Kuehl LK, Dettling M, Neuhaus AH. Assessing long-term test-retest reliability of the CPT-IP in schizophrenia. *PLoS One.* 2014;9(1):e84780.
45. Pietrzak RH, Snyder PJ, Jackson CE, Olver J, Norman T, Piskulic D, Maruff P. Stability of cognitive impairment in chronic schizophrenia over brief and intermediate re-test intervals. *Hum Psychopharmacol.* 2009;24(2):113-121.
46. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil.* 2005;84(9):719-723.
47. Jette AM, Tao W, Norweg A, Haley S. Interpreting rehabilitation outcome measurements. *J Rehabil Med.* 2007;39(8):585-590.
48. Flansbjer UB, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. *J Rehabil Med.* 2005;37(2):75-82.
49. Brown L, Sherbenou RJ, Johnsen SK. *Test of Nonverbal Intelligence: A language-free measure of cognitive ability.* Austin, TX: Pro-ed; 1990.
50. Martin JD, Blair GE, Bledsoe JR. Measures of concurrent validity and alternate-form reliability of the Test of Nonverbal Intelligence. *Psychol Rep.* 1990;66(2):503-508.
51. Stockwell DRB, Peterson AT. Effects of sample size on accuracy of species distribution models. *Ecol Modell.* 2002;148(1):1-13.

## Tables

**Table 1:** Demographic characteristics of the participants (n=103).

Characteristic	Value
Gender, n (%)	
Male	52 (50.5%)
Female	51 (49.5%)
Age range, years	
Age, mean (SD)	46.7 (10.2)
20–39	28 (27.2%)
40–49	30 (29.1%)
50–70	45 (43.7%)
Age of onset, years, mean (SD)	26.5 (8.3)
Duration of illness, years, mean (SD)	20.2 (9.3)
Education level, n (%)	
Illiterate	0
Elementary	7 (6.8%)
Middle school	38 (36.9%)
High school or vocational school	41 (39.8%)
University or graduate school	17 (16.5%)
Marital status, n (%)	
Married	28 (27.2%)
single	71 (68.9%)
Other	4 (3.9%)
CGI-S (SD)	3 (0.8)
MoCA (SD)*	23.3 (4.6)
T-SDMT (SD)*	32.1 (9.1)

\*There were some missing data in the variables of the MoCA and T-SDMT.

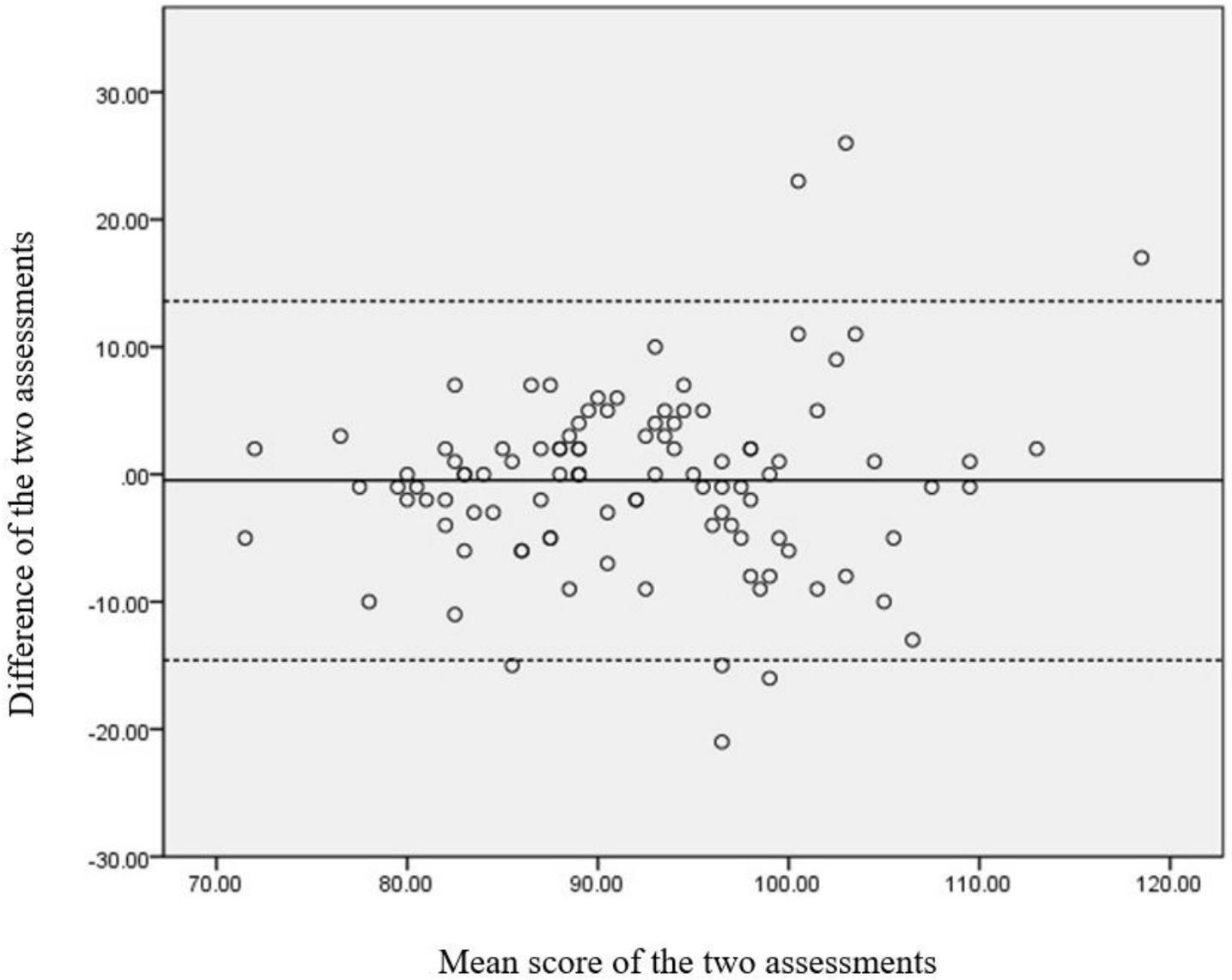
Note: CGI-S = Clinical Global Impressions Scale-Severity; MoCA= Montreal Cognitive Assessment; T-SDMT = Tablet-Based Symbol Digit Modalities Test.

**Table 2.** The mean, SD, ICC, Cohen’s d, SEM and MDC of the TONI-4 (n=103).

TONI-4	Early assessment Mean (SD)	Late assessment Mean (SD)	Cohen's d	ICC (95% CI)	SEM (SEM%)	MDC (MDC%)
Sum score	92.4 (9.1)	91.9 (9.9)	-0.03	0.73 (0.62–0.81)	4.7 (5.1)	13.1 (14.2)
Age band						
20–39	92.5 (8.7)	93.8 (10.4)	0.13	0.70 (0.45–0.85)	4.7 (5.1)	13.2 (14.2)
40–49	92.2 (9.6)	89.3 (9.7)	-0.29	0.64 (0.37–0.81)	5.8 (6.3)	16.0 (17.3)
50–70	92.4 (9.2)	92.5 (9.6)	0.00	0.82 (0.69–0.90)	3.9 (4.2)	10.9 (11.7)
Gender						
Male	92.3 (9.3)	93.4 (9.7)	0.12	0.70 (0.53–0.81)	5.1 (5.5)	14.1 (15.3)
Female	92.4 (8.9)	90.4 (10.0)	-0.21	0.76 (0.61–0.86)	4.4 (4.7)	12.1 (13.1)

Note: SD = standard deviation; ICC = intra-class correlation coefficient; SEM = standard error of measurement; SEM% = percentage of standard error of measurement; MDC = minimal detectable change; MDC% = percentage of minimal detectable change

## Figures



**Figure 1**

Bland–Altman plot showing the difference scores against the mean scores of pair of scores on the TONI-4. Note. The solid line shows the mean of the differences (-0.5). The two dashed lines represent 95% limits of agreement ( $d \pm 1.96 \times SD = -14.6-13.6$ ).