

Rating Inter-rater Reliability of Shih–Hsu Test of Attention between an Experienced Psychiatric Occupational Therapist and an Occupational Therapy Student: A Pilot Study

Yi-Nuo Shih, Ph.D.^{1,2}, Chia-Chun Wu, Ph.D.^{1,3}, Pei-Yun Shih, B.S.¹, Jia-Lien Hsu, Ph.D.⁴, Yin-Huang Liao, M.Ed.^{5*}

¹Department of Occupational Therapy, College of Medicine, Fu Jen Catholic University, ²Division of Psychiatry, Fu Jen Catholic University Hospital, ³Department of Computer Science and Information Engineering, College of Science and Engineering, Fu Jen Catholic University, ⁴Division of Psychiatry, Cathay General Hospital, New Taipei City, ⁵Psychiatric Research Center, Wan Fang Medical Center, Taipei Medical University, Taipei, Taiwan

Abstract

Objective: The “Shih-Hsu Test of Attention” (SHTA) is an iPad-based attention assessment tool developed by occupational therapists in recent years, and has acceptable criterion-related validity and high test–retest reliability in preliminary application. In this study, we intended to explore the inter-rater reliability of SHTA between experienced and inexperienced occupational therapists. **Methods:** We recruited 24 voluntary study participants aged 20–24 years in this study. The participants completed twice the SHTA by an experienced occupational therapist and an occupational therapy student. **Results:** Analytical results showed that the inter-rater reliability between experienced and inexperienced occupational therapists using SHTA had satisfactory reliability (intraclass correlation coefficient = 0.65). **Conclusion:** Our preliminary findings showed that the new attention assessment tool, SHTA, had satisfactory inter-rater reliability between experienced and occupational therapy students. We need to wait for a future study with more numbers of study participants in both groups to strengthen the study finding of this pilot study. At this moment, we suggest improving the guidance and training for inexperienced occupational therapists to improve accuracy, and reducing the gap when testing with experienced occupational therapists in future.

Key words: auditory attention, musical stimuli, occupational therapy trainees, psychiatric occupational therapy
Taiwanese Journal of Psychiatry (Taipei) 2022; 36: 138-141

Introduction

Attention performance is an important assessment item for psychiatric occupational therapy, because it remarkably affects daily life and work. Among healthy people, attention performance is important for everyday activities that requires using modern technologies such as mobile phones, computers, and machines [1, 2], and has a major effect on daily life, such as the safe operation of tools and work behavior [2].

iPad-based tools are now popular due to the rising popularity of iPad devices, and therefore, new iPad-based assessment tools need to be developed to consider convenience [3]. The “Shih-Hsu Test of Attention” (SHTA) is an iPad-based

attention assessment tool with musical stimuli developed by occupational therapists for reaction time and sustained attention. Unlike previous visual attention tests which use mostly pen and paper as the medium interface, the auditory attention assessment tool developed for this research project is used an iPad computer.

The software of iPad SHTA randomly plays several sounds, such as those from a drum, zither, piano, violin, Chinese flute, and trumpet, in 10 minutes. All sounds have the same volume

*Correspondent author. No. 280, Section 4, Renai Road, Taipei, Taiwan.
E-mail: Yin-Huang Liao <joy46373@gmail.com>

Received: Jul. 25, 2022 revised: Jul. 30, 2022 accepted: Jul. 31, 2022
date published: Sep. 28, 2022

Access this article online

Quick Response Code:



Website:
www.e-tjp.org

DOI:
10.4103/TPSY.TPSY_26_22

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: Shih YN, Wu CC, Shih PY, Hsu JL, Liao YH. Rating inter-rater reliability of Shih–Hsu Test of attention between an experienced psychiatric occupational therapist and an occupational therapy student: A pilot study. *Taiwan J Psychiatry* 2022;36:138-41.

© 2022 *Taiwanese Journal of Psychiatry* (Taipei) | Published by Wolters Kluwer - Medknow

and pitch but differ in timbre. Each sound appears randomly at intervals of 0.5, 1, and 1.5 seconds. Upon hearing each sound, a test participant is asked to press the corresponding button for “yes” or “no” on the iPad touchscreen. Scoring is based on the rate of correct answers during those 10 minutes. The highest score is 100 points; the lowest is 0 points [3, 4].

Inter-rater reliability between experienced and inexperienced clinic staff is a worthy issue for assessment tool research [5]. Inter-rater reliability is the degree of agreement among independent clinic staff who rate and assess the same assessment tool. In other words, intra-rater reliability is a score of the consistency in ratings given by the same person across multiple instances. The common way of performing reliability testing is to use the intraclass correlation coefficient (ICC). The range of the ICC may be between 0.0 and 1.0. A study explored the inter-rater reliability of Koo’s DISE classification system in the hands of experienced and inexperienced otolaryngologists, indicating that the participants’ level of experience has a strong impact on scoring, the less-experienced otolaryngologists tend to overlook some points [5]. A study indicated that moderate agreement has been observed between experienced and inexperienced examiners (ICC = 0.46) [6]. But assessments by a physiotherapist have stronger relationships to lower-limb kinematics and are more sensitive to hip joint motion than those by student assessments [7]. A study showed that visual assessment is highly consistent with the results obtained using quantitative analysis and that a substantial inter-rater agreement exists between experienced and inexperienced raters [8].

In a past study, the SHTA results have shown acceptable criterion-related validity ($\gamma = 0.400, p < 0.05$) and high test–retest reliability ($\gamma = 0.400, p < 0.05$) in healthy old adults aged 65–85 years [3]. In another study indicated that the SHTA has satisfactory test–retest reliability (ICC = 0.67), and criterion-related validity ($\gamma = 0.29, p < 0.05$) for patients with schizophrenia aged 20–64 years, that high test–retest reliability (ICC = 0.90) and criterion-related validity ($\gamma = 0.25, p < 0.05$) exist for healthy people aged 20–64 years, as well as that the value of the mapped diagnostic context percentage is 12.1%, indicating acceptable random measurement error [4].

But the SHTA has not been explored for the inter-rater reliability between experienced and inexperienced occupational therapists. A useful assessment tool needs to have reliability between experienced and inexperienced clinic staff. Therefore, in the present study, we intended to explore the inter-rater reliability between experienced and inexperienced occupational therapists using SHTA.

Methods

Study participants

This investigation studied the inter-rater reliability of the SHTA. With 24 study participants as a convenience sampling, we analyzed the ICC.

The institutional review board of Fu Jen Catholic University in New Taipei City approved the study (IRB protocol number

= C109076, and date of approval = March 12, 2021), requiring a written consent being obtained before the study. All survey and auditory data were collected anonymously. The study participants were reminded that they can withdraw from the study at any time, and that researchers have no financial interest in the study.

The study was carried out in New Taipei City from March 21, 2021, to May 24, 2021. We recruited 24 study participants aged 20–24 years for the study in New Taipei City. People with hearing impairments were excluded from this study. We enrolled 24 voluntary participants who provided informed. Two occupational therapists participated in the study: one experienced occupational therapist with >10 years of clinical psychiatry experience and one occupational therapy student.

Study procedures

Past SHTA studies used 3-week intervals when addressing test–retest reliability [3, 4]. In this study, we also used 3-week intervals.

- The SHTA was run to test the 24 voluntary participants by an occupational therapy student.
- The SHTA was run again after 3 weeks to test the same 24 voluntary participants by an experienced occupational therapist.
- The two test scores were analyzed using the ICC with a statistical software.

The measure

The SHTA is an iPad-based assessment tool that has been developed by occupational therapists to test response time and sustained attention with musical stimuli [3]. An official test takes 10 minutes, following 15 seconds of practice [3, 4].

Statistical analysis

The ICC was adopted to estimate the test–retest reliability by comparing the test scores from the two SHTA (test–retest reliability). The ICC was computed through a random effect, using two-way analysis of variance. $ICC \geq 0.90$ indicates excellent reliability; $0.75 \leq ICC < 0.90$ indicates good reliability, $0.5 \leq ICC \leq 0.74$ indicates moderate reliability, and $ICC < 0.49$ indicates poor reliability [9, 10].

We did statistical analyses for the study variable with the Statistical Package of the Social Sciences version 20 for Windows (International Business Machine SPSS Inc., Armonk, New York, USA). The differences between groups were considered significant if p values were smaller than 0.05.

Table 1. Rating intraclass correlation coefficient to compare test scores of 24 study participants from the two Shih–Hsu Tests of Attention

	Mean \pm SD	ICC
Occupational therapy student	96.250 \pm 6.257	0.650***
Experienced occupational therapist	95.833 \pm 7.340	

*** $p < 0.001$. SD, standard deviation; ICC, intraclass correlation coefficient

Results

The ICC was calculated to compare the test scores from the two SHTA between an experienced occupational therapist and an occupational therapy student.

The experienced occupational therapist was a 35-year-old female, who had been an occupational therapist for 12 years at the time of this study. She is currently a staff at a university hospital. Her rating for the scores of SHTA from 24 study participants in this study was 95.833 ± 7.340 .

The occupational therapy student was a 21-year-old female from the department of occupational therapy of a university. She had been a clinical occupational therapy student for 4 months at the time of the study. Her rating for the scores of SHTA from 24 study participants in this study was 96.250 ± 6.257 .

As shown in Table 1, the ICC of the two tests was 0.650 ($p < 0.001$).

Discussion

Attention assessment is an important issue for psychiatric occupational therapy, because attention performance substantially influences people's occupational performance and activity of daily life [2]. The development and construction of reliable and valid attention assessment tools are essential to the continued development and improvement of occupational therapy [3]. The SHTA is a recent rare auditory attention test [3, 4]. In this study, we did a preliminary examination of inter-rater reliability between experienced occupational therapists and occupational therapy students using SHTA.

In this study, the ICC of the SHTA was 0.650 between experienced occupational therapists and occupational therapy students. This study finding indicates moderate reliability [9, 10]. A past study on SHTA was done by experienced occupational therapists, and has demonstrated acceptable criterion-related validity and high test-retest reliability (ICC = 0.920) in healthy old adults aged 65–85 years [3]. The results of another study indicated that the SHTA has satisfactory test-retest reliability (ICC = 0.67) and criterion-related validity for patients with schizophrenia aged 20–64 years [4]. That study has high test-retest reliability (ICC = 0.90) and criterion-related validity for healthy people aged 20–64 years [4]. The experience of occupational therapy students is inferior to that of experienced occupational therapists, which may be the reason for the only moderate inter-rater reliability in this study.

Study limitations

The readers are advised not to overinterpret the study result, because this study still has four limitations:

- Only 24 participants took the tests in this study. The sample size is small.
- This study was to explore the inter-rater reliability of SHTA between only one experienced psychiatric occupational therapist and one occupational therapy student. The representation of the study to judge study participants'

performance with one person in each group is doubtful. The potential risk of having type 1 error exists, through failure to reject a false positive. In future studies, we need to increase the number of both groups (experienced occupational therapists and occupational therapy students) to strengthen the finding of ICC relation.

- These study participants were recruited on the basis of a convenience sampling. The representations of participants are also doubtful. Therefore, the findings may not be generalizable to other populations.
- In this study, we did not consider study participants' factors, such as education, gender, and age.

Summary

The "SHTA" is a novel iPad-based attention assessment tool with auditory stimuli, developed recently by psychiatric occupational therapists. The rating of ICC of SHTA between an experienced occupational therapist and an occupational therapy student had moderate reliability. We need to increase the numbers of both groups to strengthen the study finding of this pilot study.

Based on the findings of this study, we have two suggestions: First, we need to enhance the credibility of the inter-rater reliability, so that SHTA test should be done with a much larger sample size in future. Second, we also need to improve the guidance and training for inexperienced occupational therapists to improve accuracy and reduce the gap when testing with experienced occupational therapists in future.

Financial Support and Sponsorship

This research was supported by Grant 109-CGH-FJU-09 and Grant 110-CGH-FJU-12 from Cathay General Hospital.

Conflicts of Interest

Yi-Nuo Shih, a multidisciplinary advisory board member at the *Taiwanese Journal of Psychiatry* (Taipei), had no rôle in the peer review process or decision to publish this article. The other authors declare no potential conflicts of interest in writing this report.

References

1. Hoonakker M, Doignon-Camus N, Bonnefond A: Sustaining attention to simple visual tasks: a central deficit in schizophrenia? A systematic review. *Ann N Y Acad Sci* 2017; 1408: 32-45.
2. Shih YN, Chen CS, Chiang HY, et al.: Influence of background music on work attention in clients with chronic schizophrenia. *Work* 2015; 51: 153-8.
3. Shih YN, Hsu JL, Wu CC, et al.: Development of an iPad-based assessment tool for measuring attention and validation in older employees. *Work* 2020; 67: 811-5.
4. Shih YN, Hsu JL, Wang YC, et al.: Test-retest reliability and criterion-related validity of Shih-Hsu Test of Attention (SHTA) between people with and without schizophrenia. *Br J Occup Ther* 2022; 85: 23-8.
5. Koo SK, Lee SH, Koh TK, et al.: Inter-rater reliability between experienced and inexperienced otolaryngologists using Koo's drug-induced sleep endoscopy classification system. *Eur Arch Otorhinolaryngol* 2019; 276: 1525-31.
6. Hofauer B, Mansour N, Heiser C, et al.: Reproducibility of acoustic

- radiation force impulse imaging in thyroid and salivary glands with experienced and inexperienced examiners. *Ultrasound Med Biol* 2016; 42: 2545-52.
7. Weeks BK, Carty CP, Horan SA: Kinematic predictors of single-leg squat performance: a comparison of experienced physiotherapists and student physiotherapists. *BMC Musculoskelet Disord* 2012; 13: 207.
 8. Kahraman D, Eggers C, Holstein A, et al.: 123I-FP-CIT SPECT imaging of the dopaminergic state: visual assessment of dopaminergic degeneration patterns reflects quantitative 2D operator-dependent and 3D operator-independent techniques. *Nuklearmedizin* 2012; 51: 244-51.
 9. McGraw KO, Wong SP: Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1996; 1: 30-46.
 10. Shrout PE, Fleiss JL: Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-8.