

ORIGINAL ARTICLE

Reliability of the Modified Nine Hole Peg Test in Healthy Adults and Individuals with Hemiparetic Stroke

Narumi Watanabe, OTR, MS ^{a,b} Yohei Otaka, MD, PhD ^{a,c} Masashi Kumagai, OTR, PhD ^{a,b}
Kunitsugu Kondo, MD, PhD ^a and Eiji Shimizu, MD, PhD ^b

Objectives: This study examined whether the reliability of the Nine Hole Peg Test (NHPT) is improved by a modification (mNHPT) that confines the peg insertion/removal order to one way to reduce the degree of freedom of spatial strategies. **Methods:** Participants performed the NHPT and mNHPT three times each in two sessions with an interval of 3–5 days. Healthy adults used their non-dominant hand (n=40), while those with hemiparetic stroke used their affected (n=40) or unaffected hand (n=40). The mean value of three trials from each session was used for analyses. The reliabilities of the NHPT and mNHPT during the two sessions were assessed via intraclass correlation coefficients (ICCs) and Bland–Altman analysis. **Results:** The ICCs of the NHPT and mNHPT were 0.49 and 0.66, respectively, in healthy participants, and 0.91 and 0.94, respectively, in participants with stroke, regardless of the hand used. A significant fixed bias between the sessions was observed in both tests, except for participants with stroke who used their affected hand. Proportional biases were noted in the mNHPT results of healthy participants and in the NHPT and mNHPT results of participants with stroke who used their affected hand. The limits of agreement (lower, upper) in the affected hand were –11.0 and 9.5 for the NHPT and –8.0 and 6.2 for the mNHPT. **Conclusions:** Reduced degrees of freedom in the spatial strategy improved the relative reliability and reduced measurement errors in the NHPT. However, fixed and proportional biases were still evident.

Key Words: dexterity; fingers; motor skills; outcome assessment; validation study

INTRODUCTION

Finger dexterity is a unique characteristic of human beings and is essential for accomplishing various tasks in daily life and occupations. Around 60% of patients with middle cerebral artery stroke reportedly have some residual impairment in finger dexterity 6 months after the onset of the stroke.¹⁾ Decreased upper limb function is also reportedly associated with a decreased quality of life.²⁾ The assessment of dexterity is necessary when planning rehabilitation and evaluating the efficacy of treatment in individuals with upper extremity impairments. The Nine Hole Peg Test (NHPT) is widely used to assess finger dexterity in clinical settings because of its

simplicity, low cost, and short time to administer.³⁾ It is one of the most frequently used upper limb outcome measures in stroke rehabilitation studies.⁴⁾ When a measure is used repeatedly over time in a clinical or research setting, test–retest reliability should be considered. There are two types of test–retest reliability: relative reliability and absolute reliability. Correlation coefficients and intraclass correlation coefficients (ICCs) are commonly used to examine relative reliability. Absolute reliability is examined using the Bland–Altman analysis,⁵⁾ which systematically assesses biases and errors.

Relative reliability of the NHPT, measured using Pearson's correlation coefficient, is reportedly relatively high for the

Received: May 13, 2022, Accepted: August 16, 2022, Published online: September 9, 2022

^a Department of Rehabilitation Medicine, Tokyo Bay Rehabilitation Hospital, Chiba, Japan

^b Department of Cognitive Behavioral Physiology, Graduate School of Medicine, Chiba University, Chiba, Japan

^c Department of Rehabilitation Medicine I, School of Medicine, Fujita Health University, Aichi, Japan

Correspondence: Yohei Otaka, MD, PhD, 1-98 Dengakugakubo, Kutsukake, Toyoake, Aichi 470-1192, Japan, E-mail: otaka119@mac.com

Copyright © 2022 The Japanese Association of Rehabilitation Medicine



This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND) 4.0 License. <http://creativecommons.org/licenses/by-nc-nd/4.0/>

right hand ($r=0.69$) and moderate for the left hand in healthy adults ($r=0.43$).³ The relative reliability of the NHPT in individuals with stroke, measured using ICC, is reportedly high for both the unaffected hand ($ICC=0.89$) and the affected hand ($ICC=0.85$).⁶ Regarding the absolute reliability of the NHPT, different studies have reported the minimal detectable change (MDC) percentage in the unaffected and affected hands in stroke patients as 23% and 54%,⁶ and 12% and 24%, respectively,⁷ indicating that the measurement error may be large in some populations. Additionally, the time required to complete the NHPT has been shown to be decreased during the retest session compared to the first session in healthy adults, indicating the existence of bias.^{3,7,8} Therefore, previous studies suggest that while the relative reliability of the NHPT is acceptable, the absolute reliability is poor in terms of errors and bias.

The NHPT uses a square board with nine holes, arranged in a 3×3 square pattern, and nine pegs. The time required to place the pegs into the holes and to remove them from the holes using a single arm is measured. The order of insertion and removal of the pegs is not specified, and there are more than 360,000 ways to insert the pegs.⁹ Therefore, it is possible that the time required to complete the NHPT is influenced by the order selected by the individual. In fact, the strategy of peg insertion has been reported to influence the performance of healthy adults and individuals with stroke.⁹ When different strategies for insertion and removal of the pegs are used in the test and retest sessions, measurement errors may increase. If a more efficient spatial strategy is used in the retest session, it leads to a bias toward a shorter time to complete the test.

We hypothesized that the test–retest reliability of the NHPT would improve if a modified version (mNHPT) was used with a specific order of insertion and removal of the pegs. The purposes of this study were to examine the relative and absolute reliability of the mNHPT in healthy adults and individuals with hemiparetic stroke, and to compare them with those of the NHPT.

MATERIALS AND METHODS

Design and Setting

This study was a test–retest reliability study conducted in a convalescent rehabilitation hospital in Japan.

Participants

A total of 120 individuals participated in this study, including 40 healthy adults (mean [standard deviation, SD]

age: 26.0 [2.0] years) and 80 patients with hemiparetic stroke (mean [SD] age: 66.0 [13.3] years). Patients with their first-ever hemiparetic stroke were recruited from among those who were admitted to the convalescent rehabilitation ward via convenience sampling. The following inclusion criteria were used: patient experienced a stroke at least 1 month prior to the start of the study; patient was able to sit without assistance; patient could understand the instructions for the tasks. The following exclusion criteria were used: neurological diseases other than stroke; subarachnoid hemorrhage; lesions in the cerebellum; multiple brain lesions. Half the patients with hemiparetic stroke ($n=40$) performed the tasks with the affected hand (affected hand group), while the other half ($n=40$) performed them with the unaffected hand (unaffected hand group). Participants with hemiparetic stroke who could complete the task within 60 seconds were allocated to the affected hand group. Participants with hemiparetic stroke who could not complete the task within 60 seconds were allocated to the unaffected hand group. Healthy adults with pain and/or neuromuscular disorders in the upper extremities, cognitive deficits, or visual disturbances that affected the performance of the task were excluded. The sample size was calculated based on ICCs to ensure that the number of participants in this study was sufficient, considering a statistical power of 80% and a significance level α of 0.05. The sample size was calculated to be 36 (minimum acceptable ICC, 0.85; expected ICC, 0.95; number of repetitions, 2; expected dropout rate, 30%) based on the Sample Size Calculator (<http://wnarifin.github.io>) and methodology described by Walter et al.¹⁰

Each participant performed the test with only one arm throughout the study because inter-limb skill transfer was reported in the NHPT.⁹ The healthy adults performed the tasks with their non-dominant hand. According to the Edinburgh handedness inventory,¹¹ all but one of the healthy adults were right-handed (mean [SD] laterality quotient: 91.0 [31.9]). The characteristics of the individuals with hemiparetic stroke, including the Fugl-Meyer Assessment¹² and the modified Ashworth scale,¹³ are presented in **Table 1**. This study was conducted in accordance with the Declaration of Helsinki of 1964, as revised in 2013. The study protocol was approved by the Institutional Review Board of Tokyo Bay Rehabilitation Hospital (approval number 115-2). Written informed consent was obtained from all participants included in the study.

NHPT

The square board used in the NHPT has nine pegs and nine holes arranged in a 3×3 square pattern, spaced 3.2 cm

Table 1. Participant characteristics

Characteristic	Healthy adults (n=40)	Unaffected hand group (n=40)	Affected hand group (n=40)
Age, years	26.0 (2.0)	63.3 (14.9)	68.7 (11.0)
Sex, female/male	19/11	16/24	23/17
Type of stroke, infarction/hemorrhage	-	21/19	14/26
Side of paresis, right/left	-	19/21	26/14
Time since stroke onset, days	-	104.9 (43.6)	72.1 (28.3)
Edinburgh Handedness Inventory	91.0 (31.9)	89.8 (42.0)	95.7 (9.6)
Fugl–Meyer Assessment in the affected arm	-	18.5 (4–39)	61.5 (58–63)
Modified Ashworth scale in the affected arm ^a	-	1 (1–2)	0 (0–0)

Data are given as mean (SD), number, or median (interquartile range).

^a 1+ was treated as 2 for the modified Ashworth scale. The scores for analyses range from 0 to 5.

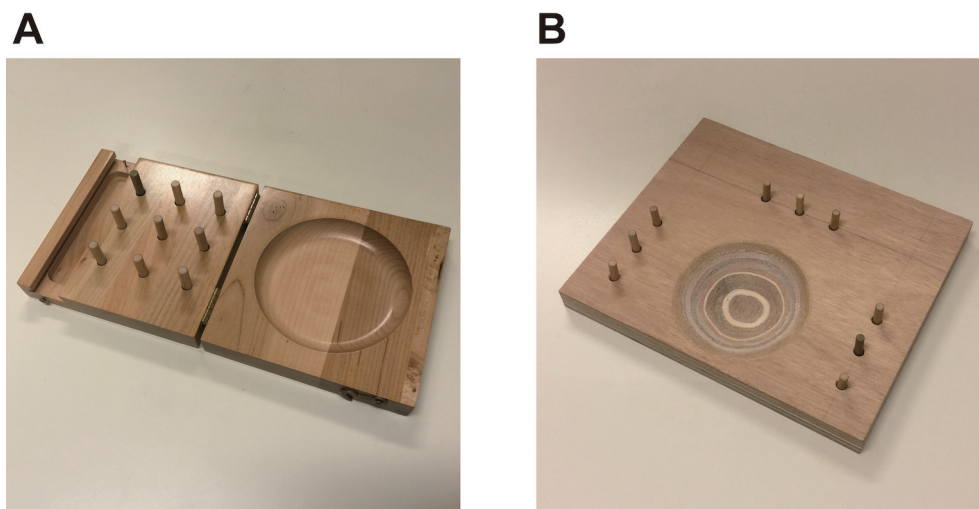


Fig. 1. Nine Hole Peg Test (NHPT) and the modified Nine Hole Peg Test (mNHPT). (A) The NHPT includes a square board with nine holes arranged in a 3×3 square and nine pegs. (B) The mNHPT confines the order of peg insertion and removal to one way to reduce the degree of freedom of spatial strategies. In both tests, the time required to place the pegs into the holes and then remove them from the holes using a single upper limb is measured.

apart when measured center-to-center (**Fig. 1A**). Each hole is 1.3 cm deep and is drilled with a 0.71-cm drill bit. The nine wooden pegs are 0.64 cm in diameter and 3.2 cm in length. The container was constructed using 0.7-cm plywood. The participants picked up the pegs and inserted them into a hole one by one using one arm until all of the holes were filled. The pegs were then removed one by one and placed in a container. The participants were instructed to perform the task as quickly as possible. The time required to complete the tasks was measured.³ If a peg was dropped outside the pegboard, the test was stopped and restarted.

Modified Version of the NHPT

We designed the mNHPT to decrease the degree of freedom in the spatial strategy for peg insertion and removal. The layout of the holes was designed such that the participant could intuitively insert or remove the pegs in one way; the layout was changed from a 3×3 square pattern to a line, and a specific order of peg insertion and removal was required (**Fig. 1B**). The spaces between the holes (3.2 cm) and the length of the pegs (3.2 cm) were the same in the NHPT and the mNHPT. The mNHPT was administered in the same manner as the NHPT, except for the peg insertion/removal order. When the participants used their left hand to complete the mNHPT, the pegs were inserted from the right lower cor-

ner to the left lower corner and removed from the left lower corner to the right lower corner.

Experimental Procedure

The NHPT and mNHPT were performed three times each during each session, and the mean duration of the three trials was used in the analyses; the second session (retest) was conducted 3–5 days after the first session.⁶⁾ The order of administration of the NHPT and mNHPT was equal among the participants to eliminate order bias. The performances of the NHPT were recorded by video, and the order of insertion and removal of the pegs was examined to evaluate the efficiency of each spatial strategy. One occupational therapist with 11 years of clinical experience supervised all tests in a quiet room throughout the study.

ANALYSES

Values for the Tests

The mean duration of the three trials in each session was calculated and used in all analyses. Differences in the values between the NHPT and the mNHPT in each session were examined using the paired *t*-test. Statistical analyses were conducted using Modified R Commander (version 4.0.2 software for Windows). *P* values <0.05 were considered statistically significant.

Relative Reliability

The test–retest reliability between the sessions was assessed using the ICC (1,1). The strength of agreement was interpreted as follows: <0.00, slight; 0–0.19, low correlation; 0.20–0.39, moderate correlation; 0.41–0.69, high correlation; 0.70–0.89, substantial; and 0.90–1.00, very high correlation.¹⁴⁾

Absolute Reliability

The absolute reliability of the NHPT and mNHPT was examined using Bland–Altman analysis⁵⁾ to check the systematic bias and estimate the limit of agreement (LOA). Two types of systemic biases exist: fixed and proportional. The two biases can be statistically confirmed and visualized using the Bland–Altman scatter plot, in which the Y-axis shows the difference between the two paired measurements, and the X-axis represents the mean of these measurements.

The fixed bias was statistically evaluated using the 95% confidence interval (CI) of the mean differences between the session 1 and session 2 values (\bar{d}). A fixed bias was present

if zero was not within the range of the 95% CI of \bar{d} . The Bland–Altman plot, in which the distribution of *d* is biased toward positive or negative, can be used to depict the fixed bias. A proportional bias was present when the value of the difference between two sessions (*d* or $|d|$) was significantly correlated with the mean of the two sessions.¹⁵⁾ The magnitude of *d* or $|d|$ changes depending on the magnitude of the mean of the two sessions in the Bland–Altman plot when a proportional bias is present.

The 95% LOA was calculated as the mean \pm 1.96 SD of the differences, and the %LOA was calculated as the mean %*d* \pm 1.96 SD of %*d*, where %*d* = (*d*/mean of sessions) \times 100, using the relative differences between sessions. These values are shown as Bland–Altman plots. In addition, the 95% CIs of the upper and lower LOAs were calculated.⁵⁾ Given that it is recommended that the acceptable LOAs be determined prior to the study,¹⁶⁾ we determined acceptable LOAs based on those calculated from previous studies conducted on individuals with stroke.^{6,7)} We calculated the LOA from the mean and SD of the differences between the sessions,⁶⁾ or from the mean and 95% CI of the differences between the sessions.^{7,17)} The calculated LOAs (lower and upper) were -7.8 and 8.6 in the unaffected hand, -24.3 and 30.5 in the affected hand;⁶⁾ and -3.3 and 1.9 in the less affected hand and -13.1 and 6.3 in the more affected hand.⁷⁾ Based on our hypothesis that mNHPT would have fewer measurement errors, we determined the acceptable LOAs in this study by calculating 80% of the mean LOAs in the two previous studies. The priori acceptable LOAs for the healthy and unaffected hands were calculated as -4.4 and 4.2 , and those for the affected hands were -15.0 and 14.7 .

In addition, the MDC score at a confidence level of 95% (MDC₉₅) was calculated using the standard error of measurement (SEM) to quantify the measurement errors: MDC₉₅ = $1.96 \text{ SEM} \times \sqrt{2}$; [MDC₉₅% = (MDC₉₅ / mean of sessions) \times 100].^{6,18)}

Strategies for Peg Insertion and Removal in the NHPT

To investigate whether the better spatial strategy improved the measurement time, we reviewed the video recordings of the NHPT trials to evaluate the efficiency of the spatial strategies of peg insertion and removal used in each trial. An efficient method of peg insertion or removal was defined as the absence of pegs that may have spatially disturbed the insertion or removal of other pegs. When the test was performed on the left arm, a point was given for inserting

or removing a peg when there was no peg in the holes of all the area on the left, lower, or lower left sides of the hole where the peg was inserted or removed (Fig. 2). The possible score ranged from 2 to 18 points for each trial, with higher scores indicating a more efficient spatial strategy. To examine whether the strategy improved during the retest, the mean score of the three trials in session 1 was compared with that of session 2 using the paired *t*-test. Pearson's correlation coefficient was used to examine the correlation between the improvement rate in the mean spatial strategy score [(session2 - session1) / mean of sessions] and that in the time required to complete the NHPT.

RESULTS

NHPT and MNHPT Values

Sessions 1 and 2 were conducted at a mean (SD) of 4.2 (0.7) days apart. The values for each session are presented in Table 2. The NHPT took significantly longer to complete than the mNHPT in each group (all *P* <0.001).

Relative Reliability

The test-retest reliability (ICC [1,1]) of the NHPT sessions was moderate in healthy adults (ICC=0.49) and very high in participants with stroke (ICC=0.91). The test-retest reliability (ICC [1,1]) of the mNHPT session was moderate in healthy adults (ICC=0.66) and very high in participants with stroke (ICC=0.94) (Table 3). The ICCs tended to be better for mNHPT than for NHPT. For both the NHPT and the mNHPT, the 95% CIs of ICCs did not overlap between the healthy adults and participants with stroke, indicating that ICCs were significantly lower in healthy adults than in participants with stroke.

Absolute Reliability

The Bland-Altman plots are shown in Figs. 3 and 4, and the data are presented in Tables 4 and 5. In all three groups,

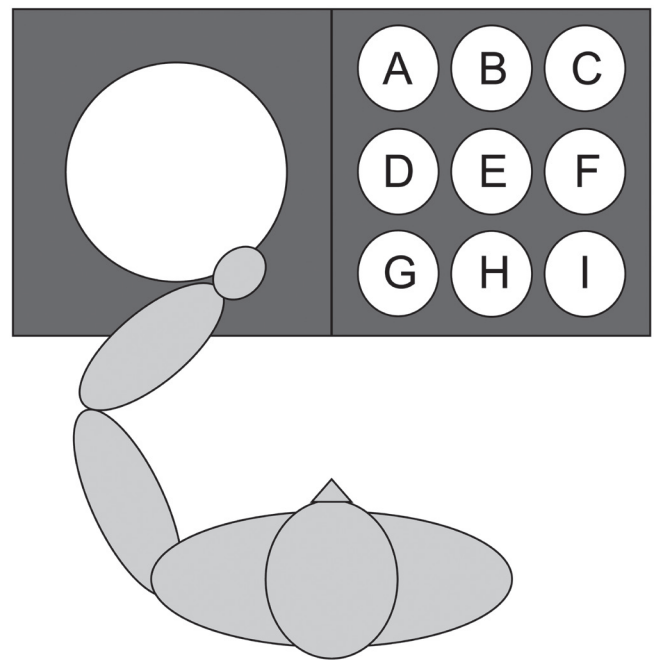


Fig. 2. Scoring the spatial peg strategy of the NHPT. The spatial strategies used by the participants to insert and remove the pegs during the NHPT were scored for efficiency. For example, when the left hand was used, the order of peg insertion/removal of CBAFEDIHG/GHIDEFABC is one of the most efficient strategies, resulting in a full score for efficiency (18 points), whereas the order of peg insertion/removal of GHIDEFABC/CBAFEDIHG is one of the least efficient strategies, resulting in the minimum score for efficiency (2 points).

the mean time required for the NHPT and mNHPT was shorter in session 2 than in session 1. The difference between each session was statistically significant in the healthy adults and the study participants with stroke who used their unaffected hand for the NHPT and the mNHPT. Therefore, a significant fixed bias was detected. In the group that used the

Table 2. Time required to complete the Nine Hole Peg Test (NHPT) and the modified Nine Hole Peg Test (mNHPT)

		Healthy adults		Participants with stroke			
		Non-dominant hand (n=40)		Unaffected hand (n=40)		Affected hand (n=40)	
		Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
NHPT	Time, s	17.6 (1.7)	16.7 (1.4)	24.2 (6.1)	23.1 (5.9)	34.7 (12.9)	33.9 (12.2)
	Strategy score	12.2 (3.7)	12.7 (3.4)	11.1 (3.6)	11.3 (3.5)	12.4 (3.3)	12.5 (3.5)
mNHPT	Time, s	16.5 (1.7)	15.8 (1.4)	22.4 (6.0)	21.5 (5.7)	31.6 (11.2)	30.7 (10.4)
P value (time of NHPT vs. time of mNHPT)		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Data given as mean (SD).

Table 3. Relative reliability of the NHPT and the mNHPT

		ICC [1, 1]	95% CI
NHPT	Healthy adults	0.49	0.22, 0.69
	Stroke (unaffected)	0.91	0.83, 0.95
	Stroke (affected)	0.91	0.84, 0.95
mNHPT	Healthy adults	0.66	0.44, 0.80
	Stroke (unaffected)	0.94	0.90, 0.97
	Stroke (affected)	0.94	0.89, 0.97

affected hand, the differences between sessions 1 and 2 were not statistically significant for the NHPT and the mNHPT, although session 2 tended to be shorter than session 1. Proportional bias was detected in healthy adults for the mNHPT and in participants with stroke who used the affected hand for the NHPT and the mNHPT (**Table 3**). Greater differences were detected when the time to complete the task was longer.

In healthy adults, the LOAs were -3.6 and 1.7 for NHPT and -2.9 and 1.5 for mNHPT. In participants with stroke who used the unaffected hand, the LOAs were -5.8 and 3.6 for NHPT and -4.3 and 2.4 for mNHPT. In participants with stroke who used the affected hand, the LOAs were -11.0 and 9.5 for NHPT and -8.0 and 6.2 for mNHPT. All LOAs except for the lower limit of the unaffected hand among the participants with stroke in NHPT were within the LOAs determined prior to the study. The LOA cannot be evaluated accurately in the presence of a proportional bias.¹⁹⁾ In this study, the LOAs of the smaller values tended to be overestimated, whereas those of the larger values tended to be underestimated in healthy adults for the mNHPT and in participants with stroke who used the affected hand in the NHPT and the mNHPT.

In healthy adults, the MDC% values of NHPT and mNHPT were 15.2% and 13.8%, respectively. In participants with stroke who used the unaffected hand, the MDC% values of NHPT and mNHPT were 20.0% and 15.3%, respectively. In the participants with stroke who used the affected hand, the MDC% values of NHPT and mNHPT were 29.9% and 22.8%, respectively (**Table 6**).

Strategies for Peg Insertion and Removal During the NHPT

No participants used the same peg strategy throughout the six trials during two sessions except for one participant with stroke who used the affected hand. The strategy scores increased slightly in the second session; however, the differences were not significant in any group (all $P > 0.05$) (**Table 2**). For either group, the rate of change in the spatial strategy

score was not significantly correlated with that of the NHPT (all $P > 0.05$).

DISCUSSION

We systematically examined the relative and absolute reliability of the NHPT in healthy adults and participants with hemiparetic stroke and examined whether the reliability was improved by modifying the test to require a specific order of peg insertion and removal, which reduces the degree of freedom of the spatial strategies. In terms of relative reliability, the ICCs in the mNHPT were better than those in the NHPT in all groups. Therefore, reducing the degree of freedom in the spatial strategy of the test improved the relative reliability. Regarding the difference between healthy adults and those with stroke, the ICCs of both NHPTs in the participants with stroke were very high (ICC, 0.91–0.94) and significantly better than those in healthy adults (ICC, 0.49–0.66). This difference between the healthy adults and those with stroke might have been caused by the difference in the range (variability) of values in the samples, because a larger ICC is obtained in a sample with a larger range. The range of values in the affected and unaffected hands in the participants with stroke was markedly larger than that in healthy adults; therefore, the ICCs might have been larger in participants with stroke than in healthy adults.

Regarding absolute reliability, the range of the upper and lower LOA in the mNHPT was narrower than that in the NHPT in all groups. Additionally, the LOA of the NHPT in the unaffected hand group was outside of the priori determined acceptable limits. Furthermore, the MDC% values were smaller in the mNHPT than in the NHPT for all groups. The MDC% in the mNHPT in this study was 22.8%, which was smaller than the MDC% in the NHPT in the affected hand in previous studies, which have been reported to be 24%⁷⁾ and 54%.⁶⁾ This suggests that the mNHPT has a smaller measurement error.

The improved relative reliability and reduced measurement error observed in the mNHPT might have been caused by the reduced variability of task performance for each participant because of the requirement of a specific spatial strategy. In fact, only one participant out of 120 (0.8%) performed the NHPT with the same order of insertions and removals throughout the trials. Therefore, it is reasonable to assume that the difference in the order of pegs between the trials increased the variability of the NHPT. Considering that there was no correlation between the spatial strategy score and the time required for the NHPT, the variability in the

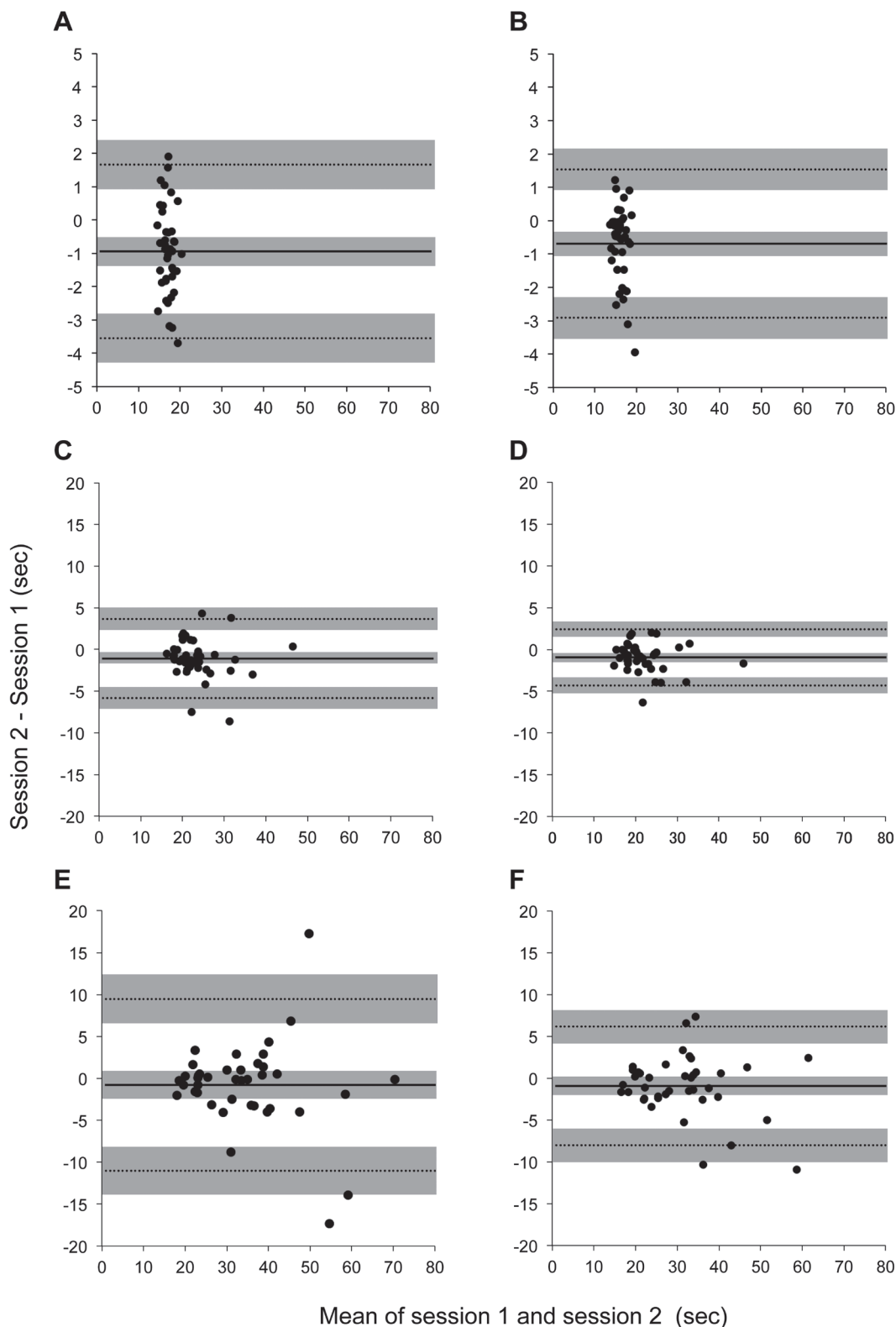


Fig. 3. Bland–Altman plots for the NHPT (A, C, E) and mNHPT (B, D, F). (A, B) Data of the trials of healthy adults using their non-dominant hands. (C, D) Data of the trials of participants with stroke using the unaffected hand. (E, F) Data of the trials of participants with stroke using the affected hands. Solid lines represent the mean and dotted lines represent the LOAs. The shaded areas represent the 95% confidence intervals for the mean and LOAs.

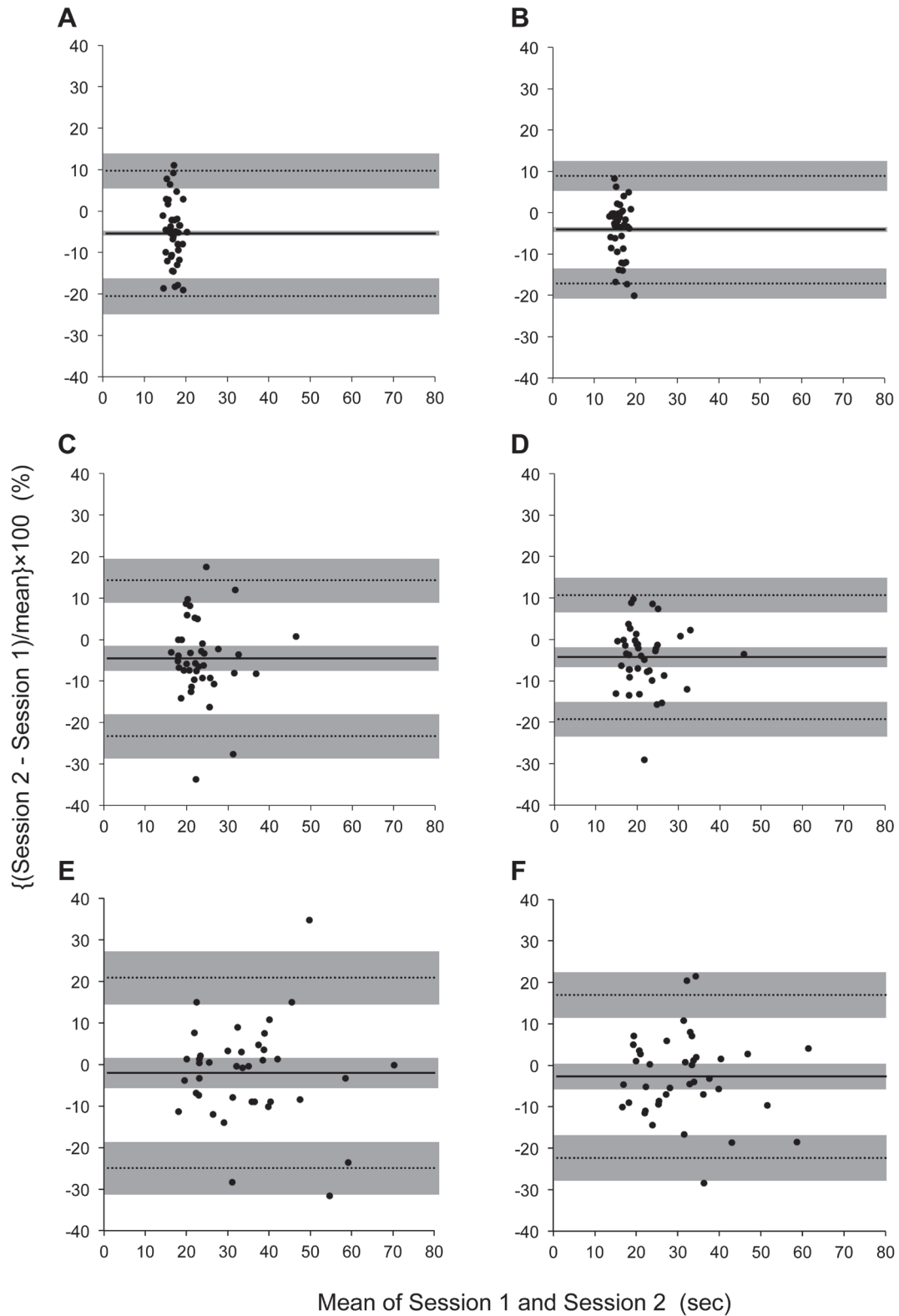


Fig. 4. Bland–Altman plots for the NHPT (A, C, E) and mNHPT (B, D, F) using the relative difference between the sessions. (A, B) Data of the trials of healthy adults using their non-dominant hands. (C, D) Data of the trials of participants with stroke using the unaffected hand. (E, F) Data of the trials of participants with stroke using the affected hands. Solid lines represent the %mean and dotted lines represent the %LOAs. Shaded areas represent 95% confidence intervals for the %mean and %LOAs.

Table 4. Absolute reliability of the NHPT and the mNHPT with Bland–Altman analysis

		Fixed bias						Proportional bias			
		\bar{d}	95% CI	Lower	95% CI	Upper	95% CI	d vs mean		d vs mean	
				LOA		LOA		r	P	r	P
NHPT	Healthy adults	-0.9	-1.37, -0.52	-3.6	-4.28, -2.82	1.7	0.94, 2.40	-0.21	0.200	0.19	0.229
	Stroke (unaffected)	-1.1	-1.86, -0.32	-5.8	-7.16, -4.49	3.6	2.30, 4.97	-0.11	0.498	0.14	0.240
	Stroke (affected)	-0.8	-2.45, 0.89	-11.0	-13.90, -8.12	9.5	6.56, 12.34	-0.15	0.367	0.49	0.001
mNHPT	Healthy adults	-0.7	-1.05, -0.33	-2.9	-3.55, -2.29	1.5	0.91, 2.17	-0.31	0.049	0.38	0.014
	Stroke (unaffected)	-1.0	-1.50, -0.40	-4.3	-5.27, -3.37	2.4	1.46, 3.36	-0.16	0.327	0.15	0.231
	Stroke (affected)	-1.0	-2.07, 0.24	-8.0	-9.99, -5.99	6.2	4.17, 8.17	-0.22	0.179	0.47	0.002

Table 5. Absolute reliability of the NHPT and the mNHPT with Bland–Altman analysis using the percentage of difference

		Fixed bias					
		%Mean difference	95% CI	Lower %LOA	95% CI	Upper %LOA	95% CI
NHPT	Healthy adults	-5.4	-5.83, -4.98	-20.6	-24.82, -16.27	9.7	5.47, 14.01
	Stroke (unaffected)	-4.5	-7.59, -1.46	-23.3	-28.62, -18.01	14.3	8.96, 19.57
	Stroke (affected)	-2.0	-5.74, 1.70	-25.0	-31.4, -18.47	20.9	14.43, 27.36
mNHPT	Healthy adults	-4.1	-4.49, -3.76	-17.1	-20.80, -13.45	8.9	5.21, 12.55
	Stroke (unaffected)	-4.3	-6.72, -1.84	-19.3	-23.48, -15.03	10.7	6.46, 14.91
	Stroke (affected)	-2.7	-5.90, 0.51	-22.4	-27.89, -16.80	17.0	11.41, 22.51

Table 6. Minimal detectable changes (MDC) for the NHPT and the mNHPT

		SEM	MDC ₉₅	MDC _{95%}
NHPT	Healthy adults	0.94	2.6	15.2
	Stroke (unaffected)	1.71	4.7	20.0
	Stroke (affected)	3.69	10.2	29.9
mNHPT	Healthy adults	0.80	2.2	13.8
	Stroke (unaffected)	1.21	3.4	15.3
	Stroke (affected)	2.56	7.1	22.8

time required for the cognitive process for identifying spatial strategies, rather than the variability in the spatial barriers of the pegs, might be responsible for the reduced reliability of the NHPT. We believe that by uniquely determining the order of pegs as in mNHPT, we may reduce not only the variability in spatial strategies, but also the variability in the time required for completion of the cognitive process for identifying the spatial strategy.

Contrary to our expectations, the reduction in the degree

of freedom of the spatial strategies did not eliminate the fixed biases observed in the healthy adults and participants with stroke who used the unaffected hand. The values in the second session were significantly smaller than those in the first session. Furthermore, there was no significant relationship between the spatial peg strategies and the time required to complete the NHPT, although various peg strategies were used by the participants. This finding suggests that the main cause of the fixed bias observed in the NHPT was not the

improvement of spatial strategy but the learning effect of peg manipulation itself. A statistically significant fixed bias was not observed in the affected hand of participants with stroke, although there was a tendency towards this bias. This may have been caused by the fact that the learning effects were obscured by the variability of task performance, or that the learning ability was impaired in the affected hand of the participants with stroke.

Proportional bias was detected in the healthy adults in the mNHPT and in the participants with stroke who used the affected hand in the NHPT and the mNHPT. This bias might have been simply caused by the variability in performances between the sessions. As the time required for the test increased, the variability between the sessions might have increased. Regarding the relative and absolute reliability, similar findings were reported in the Purdue Pegboard Test (PPT), which is known as a test for finger dexterity. The PPT was reported to have almost perfect test–retest reliability (ICCs >0.8 for each subtest in healthy adults).²⁰⁾ However, systemic bias was present, and a significantly more favorable result was observed in the second test compared to the first test in some subtests of the PPT in individuals with schizophrenia.²¹⁾

In terms of clinical implications that can be drawn from this study, the mNHPT can be used with better relative and absolute reliability in terms of reducing the measurement error compared to that of the NHPT when assessing finger dexterity in healthy adults and those with stroke. However, the fixed bias caused by the learning effect when the test is used repeatedly cannot be overlooked. Furthermore, clinicians should consider that the test has some measurement error, even though the LOAs are acceptable, and that these errors become greater because of proportion bias when the test is used in individuals with severe impairment who require more time to complete the test. Importantly, the measurement time differed between the NHPT and the mNHPT, which might be caused by the differences in test structures, indicating the risk involved in using the values obtained using the NHPT in direct comparison with those obtained using the mNHPT.

The limitations of the study include the small sample size and the lack of age-matched controls. Therefore, the results may not be generalizable to individuals with stroke with different degrees of hemiparesis. In addition, it could not be determined whether the characteristics observed in the unaffected hand were stroke-specific or because of aging, although the learning effect was evident even in participants with stroke who used the unaffected hand.

CONCLUSION

The mNHPT has better absolute and relative reliability in terms of reducing the measurement error than the NHPT in healthy adults and individuals with stroke. However, fixed bias, proportional bias, and measurement errors cannot be ignored.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

1. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJ: Probability of regaining dexterity in the flaccid upper limb: impact of severity of paresis and time since onset in acute stroke. *Stroke* 2003;34:2181–2186. DOI:10.1161/01.STR.0000087172.16305.CD, PMID:12907818
2. Franceschini M, La Porta F, Agosti M, Massucci M: Is health-related quality of life of stroke patients influenced by neurological impairments at one year after stroke? *Eur J Phys Rehabil Med* 2010;46:389–399. PMID:20927005
3. Mathiowetz V, Weber K, Kashman N, Volland G: Adult norms for the Nine Hole Peg Test of finger dexterity. *Occup Ther J Res* 1985;5:24–38. DOI:10.1177/153944928500500102
4. Santisteban L, Térémetz M, Bleton JP, Baron JC, Maier MA, Lindberg PG: Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. *PLoS One* 2016;11:e0154792. DOI:10.1371/journal.pone.0154792, PMID:27152853
5. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;327:307–310. DOI:10.1016/S0140-6736(86)90837-8, PMID:2868172
6. Chen HM, Chen CC, Hsueh IP, Huang SL, Hsieh CL: Test–retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke. *Neurorehabil Neural Repair* 2009;23:435–440. DOI:10.1177/1545968308331146, PMID:19261767
7. Ekstrand E, Lexell J, Brogårdh C: Test–retest reliability and convergent validity of three manual dexterity measures in persons with chronic stroke. *PM R* 2016;8:935–943. DOI:10.1016/j.pmrj.2016.02.014, PMID:26972364

8. Oxford Grice K, Vogel KA, Le V, Mitchell A, Muniz S, Vollmer MA: Adult norms for a commercially available Nine Hole Peg Test for finger dexterity. *Am J Occup Ther* 2003;57:570–573. DOI:10.5014/ajot.57.5.570, PMID:14527120
9. Iosa M, Morone G, Ragolini MR, Fusco A, Paolucci S: Motor strategies and bilateral transfer in sensorimotor learning of patients with subacute stroke and healthy subjects. A randomized controlled trial. *Eur J Phys Rehabil Med* 2013;49:291–299. PMID:23172404
10. Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101–110. DOI:10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E, PMID:9463853
11. Oldfield RC: The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113. DOI:10.1016/0028-3932(71)90067-4, PMID:5146491
12. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S: The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scand J Rehabil Med* 1975;7:13–31. PMID:1135616
13. Bohannon RW, Smith MB: Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther* 1987;67:206–207. DOI:10.1093/ptj/67.2.206, PMID:3809245
14. Guilford JP: *Fundamental statistics in psychology and education*. McGraw-Hill, New York and London, 1942.
15. Altman DG, Bland JM: Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983;32:307–317. DOI:10.2307/2987937
16. 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. DOI:10.1093/bja/aew320, PMID:27799171
17. Bland M: *An introduction to medical statistics*. Oxford University Press, Oxford, 2015.
18. Faber MJ, Bosscher RJ, van Wieringen PC: Clinimetric properties of the performance-oriented mobility assessment. *Phys Ther* 2006;86:944–954. DOI:10.1093/ptj/86.7.944, PMID:16813475
19. Ludbrook J: Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* 2002;29:527–536. DOI:10.1046/j.1440-1681.2002.03686.x, PMID:12060093
20. Buddenberg LA, Davis C: Test–retest reliability of the Purdue Pegboard Test. *Am J Occup Ther* 2000;54:555–558. DOI:10.5014/ajot.54.5.555, PMID:11006818
21. Lee P, Liu CH, Fan CW, Lu CP, Lu WS, Hsieh CL: The test–retest reliability and the minimal detectable change of the Purdue Pegboard Test in schizophrenia. *J Formos Med Assoc* 2013;112:332–337. DOI:10.1016/j.jfma.2012.02.023, PMID:23787010