IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 July 2023, accepted 29 August 2023, date of publication 1 September 2023, date of current version 12 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3311270

RESEARCH ARTICLE

Validity and Reliability of a New Developed Digital Version of Nine Hole Peg Test

EVELINE PROCHASKA[1,2], AND ELSKE AMMENWERTH[2]
[1]Department of Applied Electronics and Technical Informatics, University of Applied Sciences Campus Vienna, 1100 Vienna, Austria
[2]Department of Health Information Systems, Private University for Health Sciences and Health Technology, 6060 Hall in Tirol, Austria

Corresponding author: Eveline Prochaska (eveline.prochaska@fh-campuswien.ac.at)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the University of Applied Sciences Campus Vienna under EK Nr. 97/2022, and performed in line with the Declaration of Helsinki.

ABSTRACT In healthcare, the Nine Hole Peg Test (NHPT) is considered a standard for assessing hand dexterity. A digital version of the classic Nine-Hole Peg Test – the dNHPT - has been developed that allows digital measurement of hand function. The prototype of this dNHPT was investigated in this study with test-retest and crossover design with 32 healthy adults. Pearson Correlation Coefficient and Bland-Altman diagram were used to analyze concurrent validity. Intraclass Correlation Coefficients (ICC), standard error or measurement (SEM) and smallest detectable change (SDC) were used to determine test-retest and interrater reliability. Our results showed a moderate concurrent validity ($r = 0.592$). The Bland Altman analysis showed an estimated a mean difference of -2.47 between the dNHPT in comparison with the conventional NHPT. The dNHPT demonstrated good test-retest reliability (ICC: 0.75, SEM: 0.89, SDC: 2.47) and high interrater reliability (ICC: 0.76). To conclude, the dNHPT can contribute to objectify the measurement of hand dexterity without losing its most compactness and simplicity as the most important properties of NHPT.

INDEX TERMS Biomedical equipment, engineering in medicine and biology, patient rehabilitation, product development, product validation, prototypes, reliability.

I. INTRODUCTION

Assessments collect data using instruments to evaluate outcomes of therapeutic interventions [1]. For assessing hand dexterity the Nine Hole Peg Test (NHPT) is one of the most commonly used tool [2]. Dexterity is the ability to make purposeful, coordinated hand and finger movements to grasp and manipulate objects [3]. Adequate fine motor dexterity is crucial for almost all tasks in daily living [4].

Dexterity assessments often used in clinical settings to ascertain an individual´s hand function [5]. The NHPT assessment was scored on the number of seconds it took subjects to insert nine pegs into a board and then remove then. The resulting scores are compared with previous measurements, e.g. to assess a therapy effect or a change in the context of a disease [3] or compared with clinical norm data [6]. NHPT is a simple and easy tool for screening fine motor dexterity. The original NHPT does not contain any technology. It is made entirely of either wood or plastic and also requires a stopwatch to perform the standardized test procedure. Various research projects have already worked on digitizing the NHPT. Johansson et al. developed the modified NHPT using cameras and markers on the body of the person being tested [7]. Some other developments use virtual environments to assess dexterity along the example of the NHPT [8], [9], [10]. All these developments have in common that the easy handling of the conventional NHPT is lost. A considerable amount of equipment is required and therefore technical understanding from users. At the same time, data collection is automated and improved.

We have thus developed a digital version of a NHPT (dNHPT) that combines the advantages of the ease of use

The associate editor coordinating the review of this manuscript and approving it for publication was Genoveffa Tortora.

and automatic data collection. In the previous research that digitized the NHPT, a lot of additional equipment became necessary (cameras, sensors, powerful PCs, VR equipment). Our newly developed prototype consists only of the test board and an additional control unit, thus providing users with the familiar form (as with the original NHPT). However, the dNHPT includes electronics and software to improve the test procedure and data collection. In order to assess the quality of a measurement procedure (assessment) in healthcare, the following criteria are used: reliability and validity [11]. In this study, the measurement properties of the newly developed prototype (dNHPT) are evaluated.

In particular, this study aimed to evaluate (1) the concurrent validity of the dNHPT, (2) the test-retest reliability and (3) the interrater-reliability of the dNHPT.

II. MATERIAL AND METHODS

Study methods follows the standard protocol items: COSMIN (Consensus based standards for the selection of health status measurement instruments), which is the only consensus-based checklist for the preferred design characteristics and statistical methods of studies on measurement properties [12].

A. NINE HOLE PEG TEST (ORIGINAL)

The Nine Hole Peg Test (NHPT) was developed by Kellor et al. [13] which is considered as a gold standard measurement of manual dexterity [14]. The Jamar Nine Hole Peg Test (Smith & Nephew) was used. the structure is divided in two parts, a pegboard with nine recesses arranged in a grid 3 × 3 and on a flat round tray containing the nine pegs. This version differs from the original version according to Mathiowetz [6] in the shape of the container where the pegs are stored. There are no statistically significant differences between the two versions [2].

Participants have to grasp, one by one, the nine pegs from the container, inserting each one into a hole until all pegs are placed. After this the pegs have to replaced one by one back into the container. The task should be performed as fast as possible [Math]. The test was timed with a stopwatch from the moment the participant touched the first peg until the moment when the last peg was replaced into the container [6].

B. THE NEWLY DEVELOPED PROTOTYPE

We have developed the dNHPT to further standardize the measurement with the NHPT by using digital functions to verify the test procedure and automatically measure the time. Figure 1 shows an overview of the dNHPT. This prototype is able to record the time of the test procedure, store the recorded data and to monitor the correct execution. Monitoring checks whether each pin was inserted and removed properly.

1) HARDWARE

The dNHPT consists of the control unit and the pegboard. The pegboard is in form and dimensions exactly oriented to the specifications of Mathiowetz [6]. The dNHPT was designed using Inventor 2019 CAD software and produced with a 3D printer using PLA (poly lactic acid). The pegboard has nine holes with a distance of 21 mm (from hole center to hole center). The diameter of the holes is 7.1 mm and they are 13 mm deep. The pegs have a length of 32 mm and a diameter of 6.4 mm.

FIGURE 1. The components of the dNHPT (test board with control unit, peg container and voltage supply).

The prototype of the dNHPT we made is shown in Fig. 1. The container for the pegs can be used to the right or left of the test board depending on the user, magnets hold the container in a stable position. The control unit includes a 4-line display, a 4 × 4 matrix keypad, 3 buttons (start, stop and reset) and an SD card module. The supply voltage (5 V) is provided by a USB 2.0 type A, at least 400 mA are required (see Fig. 1).

2) ELECTRONICS

The pegs are equipped with neodymium magnets (Type: N42) with 2mm diameter and 1mm thickness at both ends of each peg. The magnetic flux density (B) of the neodymium magnet was calculated by the following formular (1). Br= the magnetic remanence, D = the magnet thickness, R = radius of the magnet, z = distance from the pole face on the symmetry axis.

$$B = \frac{Br}{2}\left(\frac{D+z}{\sqrt{R^2+(\mathrm{D}+\mathrm{z})R^2}} - \frac{z}{\sqrt{R^2-z^2}}\right) \quad (1)$$

The nine digital hall sensors (TLE4906L) are used with internal circuitry as shown in Figure 2. The sensors are implemented at the test board detecting the magnetic fields of the pegs when the distance between magnet and sensor is $\leq$ 3mm. The detection of the magnetic fields of the nine pegs serves monitoring for error-free execution of the test procedure and for function testing of the prototype.

The whole system is controlled by a 32-bit microcontroller STM32F4 on a discovery board. The microcontroller board is located in the control unit, the sensors for detecting the inserted pins are installed directly on the test board (see Fig. 3).

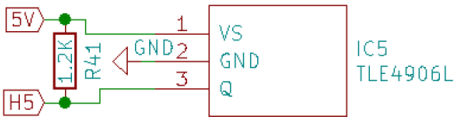

FIGURE 2. Internal circuit diagram of hall sensors.

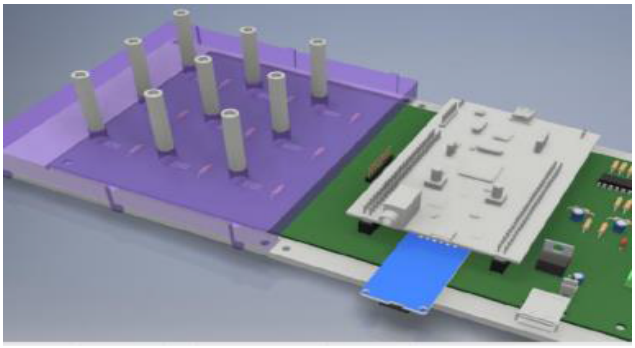

FIGURE 3. Arrangement of hall sensors, pegs and microcontroller.

3) SOFTWARE

The software was developed especially for this system in the language C. The prototype provides 3 functions: standard test, trial test and functional test (Figure 4). The standard test enables the execution of the standardized NHPT procedure. The Start button starts the time measurement, the complete insertion and removal of the pegs is tested, the Stop button ends the time measurement. A valid measurement exists when all pegs have been inserted and then removed again. The result in seconds is shown on the display and can be stored on the SD card.

The trial test - which is provided in the standardized procedure for understanding the test procedure - differs in functionality only in that no data is saved.

The function test checks the prototype for possible problems, such as missing magnets, defect of a sensor, contamination of the holes.

The prototype of the dNHPT enables the assessment according to the standardized specifications [5]. Errors during the execution, such as incorrect plugging of the pegs, are indicated on the display.

C. STUDY DESIGN

This research follows a test-retest design with crossover. The participants were randomly matched to two groups. NHPT resp. dNHPT data were collected at two measurement time points with crossover after the first measurement point. The total data collection period was ten days. Two testers (rater 1 and rater 2) conducted all data collection. Prior to the study, the two testers performed two pre-tests. The pretests were conducted once with two testers and one proband and once with only the two testers. The pretest aimed to train the testers and to check the planned procedure of the surveys in order to increase the reliability of the data to be collected.

D. PARTICIPANTS

In total, 32 persons participated in this study. Participants were students of the university of applied sciences in Vienna (Austria). The sample size calculation for evaluating the correlation was calculated with G∗Power Version 3.1.9.7. A sample size >= 30 was considered sufficient for group comparison. The inclusion criteria were (1) individuals without history of neuromuscular or orthopedic dysfunction that would significantly affect dexterity, (2) $\geq$ 18 years of age. Handedness was identified by asking the participant which hand was used for writing.

E. ETHICS AND REGISTRATION

The study protocol was in accordance with the Declaration of Helsinki and was approved by the ethics committee (EK Nr. 97/2022) of the University of Applied Sciences Campus Vienna. All subjects provided their written informed consent before the study.

This study has been registered on open science framework and the registration number is DOI 10.17605/OSF.IO/BW2M4 (registration information is available at https://osf.io/bw2m4).

F. EXPERIMENTAL PROCEDURE

The study design includes two measurement time points. The test procedures took place in a room specially prepared for this purpose at the University of Applied Sciences Campus Vienna. The setting and test instructions for the NHPT and the digital NHPT corresponded to the standard set by Mathiowetz [6]. Test instruction were translated into German by the author. One measurement of the writing hand of each participant was performed. Participants sat on a chair in front of a table. The test board was centrally located in front of them. The peg box was on the side of the hand to be tested. The instructions for the test were read out by the tester according to the standardized instructions, including a short demonstration. The participants performed a practice run (without timing) prior to the recorded test. The tests were timed, at NHPT with a stopwatch, at digital NHPT with the implemented time measurement at the push of a button. In case that the participant dropped a pen while performing the test, the test was terminated and a new one was started.

Data collection took place at two measurement times, with ten days in between. This period was chosen to be small enough so that no change in hand function occurs, but at the same time large enough to minimize influences from practice or memory [10], [15]. 32 participants were randomized into both groups, resulting in 16 participants in group 1 and 16 to group 2. At the first measurement group 1 was tested from tester 1 with the digital NHPT and then at the same day using the conventional NHPT. Group 2 was tested by tester 2 in reverse order (first conventional NHPT, and then digital NHPT).

To avoid a possible learning effect by repeated measurement performance, a measurement with another measurement

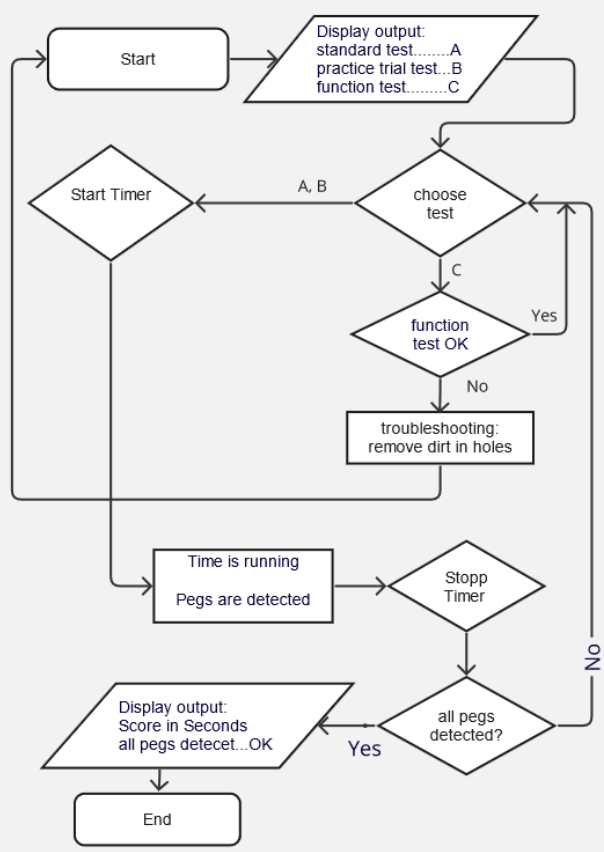

FIGURE 4. Flowchart explaining programming of the setup.

instrument, the box-and-block test, was performed between the measurements of dNHPT and NHPT. The measurements followed written, standardized instructions by the two trained testers to prevent bias.

At the second measurement point, 10 days after first measurement, total 17 participants took part. Both groups tested using dNHPT. Here, both groups changed the tester – group 1 was thus tested by tester 2 and group 2 by tester 1.

This study design was chosen to allow assessing both test-retest-reliability, interrater-reliability and validity, compared to NHPT, of dNHPT.

G. DATA ANALYSIS

We used IBM SPSS Statistics version 28.0 for data analysis. Descriptive statistics was used to describe the study population. Normality of data was evaluated using Shapiro Wilk test. Concurrent validity was determined by Pearson correlation coefficient for the relationship between the conventional NHPT and the digital NHPT at measurement point one. Following correlation classification: no or very low: ρ =0–0.25; low: ρ =0.26–0.40; moderate: ρ =0.41–0.69; high: ρ =0.70–0.89; very high: ρ =0.90–1.0 [16]. The level of statistical significance was chosen with $p \leq 0.05$.

TABLE 1. Participant characteristics (n = 32).

sex, *n (%)*	female	30 (94%)
	male	2 (6%)
age *(years)*	mean ±SD	21.5 ±6.25
tested hand	right	30 (94%)
	left	2 (6%)

The agreement between NHPT and dNHPT was examined using Bland-Altman analysis to check the systematic bias and estimate the limit of agreement (LOA) [11], [17]. In the Bland-Altman scatter plot the x-axis represents the mean of these measurements and the y-axis shows the difference between the two paired measurements. The fixed bias was statistically evaluated using the 95% confidence interval (CI) of the mean differences between the NHPT and dNHPT values at measurement point 1. A fixed bias is present when if zero is not within the range of the CI. After ensuring that the differences are normally distributed, standard deviation (SD) can be used for define the LOA: mean $\pm 1.96 * \text{SD}$ [18]. The limits of agreement show how much the scores can vary in stable patients. A change in scores in scores within LOAs or smaller indicates a measurement error, outside the LOAs it can be assumed these are statistically significant changes [11].

For assessing interrater- and test-retest reliability intraclass correlation coefficients (ICC) were investigated. To estimate the correlation following classification of correlation was used [19]: $<$ 0.5 poor, 0.5 - 0.75 moderate, 0.75 - 0.9 good and $>$ 0.9 excellent. An ICC of 0.7 or greater is an accepted minimum for reliability of measurement methods (assessments) [11]. Measurement error was determined by estimating the standard error of measurement (SEM) using the formula $SEM = SD\sqrt{(1-ICC)}$, where SD is the standard deviation of the means from all probands [11] and ICC from the test-retest reliability. Smallest Detectable Change (SDC) was calculated, based on the test-retest parameter SEM, as follows: $\text{SDC} = \text{SEM} * 1.96 * \sqrt{2}$ [11]. The SDC represents the minimal change that a patient must show on the scale to ensure that the observed change is real and not just measurement error [20].

III. RESULTS

A total of 32 healthy subjects participated in this study. Their characteristics are summarized in Table 1. The mean age of the participants was 21.5 $\pm$ 6.25 years. The majority of the participants (n = 30) were righthanded and 2 participants were lefthanded.

Table 2 shows the means, standard deviations, maximum and minimum scores, and the number of valid values of the three measurements with NHPT, dNHPT1 (both at the first measurement time point), and dNHPT2 (at the second measurement time point). The NHPT shows on average lower scores than dNHPT1 and dNHPT2, the score ranges (in seconds) are for NHPT: 7.1, dNHPT1: 9.44 and dNHPT2: 5.82.

TABLE 2. Average performance of healthy persons on NHPT and DNHPT (in seconds).

	Mean	SD	Minimum	Maximum	Valid values
NHPT	14.39	1.53	12.3	19.4	32
dNHPT1	16.81	2.05	13.15	22.59	32
dNHPT2	16.85	1.51	20.55	14.73	17

NHPT (scores of original NHPT at measurement point 1), dNHPT1 (scores of dNHPT at measurement point 1), dNHPT2 (scores of dNHPT at measurement point 2).

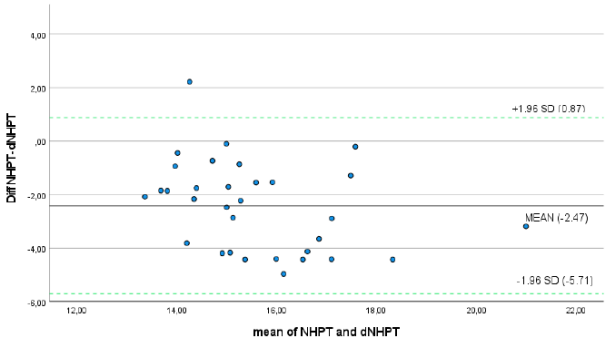

FIGURE 5. Bland-Altman-Plot of NHPT and dNHPT1.

A. CONCURRENT VALIDITY

At the first measurement point, correlations between NHPT and dNHPT were moderate ($r = 0.592$) and significant ($p < 0.001$).

The Bland-Altman limits of agreement (LOA), also named Bland-Altman plots, analyses the pairs of observations (conventional NHPT and dNHPT) from the same subjects (shown in Figure 3). The means and differences of these pairs of values for each subject are displayed in a scatter plot. The plot shows also a line for the estimated mean difference between the two versions of the NHPT with -2.47 [95% CI: -1.84; 3.01]. The two dashed lines indicating the Limits of Agreement (LOA: +1.96SD: 0.87; -1.96 SD: -5.71).

The LOAs give an indication of how much the scores can vary in stable probands. The mean difference is -2.47 what means, that on average the dNHPT measures 2.47 seconds more than the NHPT. Results measured by NHPT may be −5.71 seconds below or 0.87 seconds above dNHPT scores. The 95% CI of the mean difference did not include the line of equality, so a consistent bias was found. The Bland Alman plot indicates that all but one collected score is within the limits of agreement [0.87; -5.71].

B. TEST-RETEST RELIABILITY

For the calculation, the scores of the dNHPT (execution time in seconds) were compared at the two measurement points. From the whole sample with 32 healthy participants, 17 also completed the second measurement point 2. Test-retest reliability was determined through calculating the ICC (3, k), based on the 2-way mixed (k fixed raters are defined), absolute agreement (agreement between two raters is of interest), average measure [11], [21]. The intraclass correlation coefficient is ICC = 0.75 [0.28; 0.91] and significant ($p < 0.05$).

Since the ICC is only an expected value of the true ICC, it is appropriate to assess the degree of reliability on the basis of the 95% confidence interval of the ICC value and not the ICC value itself [19]. This results in an interpretation of the test-retest reliability level of the dNHPT from poor to good. The SEM was 0.89, indicating that the dNHPT is capable of reasonable estimates of patient performance. The result of SDC = 2.47 in healthy adults' states that we can assume that 95% of the tested population has a random variation of less than 2.47 seconds on repeated testing. A value above 2.47 would indicate a true change (beyond an expected measurement error).

C. INTERRATER RELIABILITY

Interrater reliability was assessed with ICC (2, k) since, in contrast to test-retest reliability, the focus here is on the context of repeated measurements of the same participants by two raters [21]. For this purpose, the results of tester 1 and tester 2 were compared for the 17 participants that completed dNHPT at both measurement points. Interrater reliability was significant ($p < 0.05$), with ICC = 0.76 [0.33; 0.91].

IV. DISCUSSION

The aim of this study was to evaluate the concurrent validity, the test-retest reliability and the interrater reliability of the newly developed dNHPT. The concurrent validity of dNHPT was assessed in comparison with the conventional NHPT, which is considered as a gold standard measurement of manual dexterity [14]. With test repetitions at two measurement time points and the use of 2 testers the reliability was evaluated. This paper presented results of a study with 32 healthy participants performing repetitions of the digital Nine Hole Peg Test.

Other papers have presented various conventional NHPT assessments, such as the original NHPT in addition to further technologies [7], alternative devices instead of the NHPT [8] and NHPT additionally using virtual reality [9], [10], [22], [23]. Compared to other studies, a unique feature of the dNHPT is that no additional, technical equipment is required. It is a stand-alone solution, like the conventional NHPT. At the same time, it provides digital functions that support the execution of the measurement (guidance through the measurement process, time measurement and control for error-free execution).

A. CONCURRENT VALIDITY OF THE DNHPT

The comparison of the new dNHPT with the gold standard NHPT has resulted in a moderate correlation. However, a correlation coefficient is highly dependent on the variability of the sample. Since our sample has a low variability, as it was only healthy subjects, the correlation coefficient alone is not meaningful enough. The correlation coefficient alone

is not sufficient to assess the agreement of two measurement methods [17], so in a further step the Bland Altman analysis was performed. The Bland Altman plot found that the dNHPT scores were in average 2.47 seconds (14.4%) higher than the NHPT scores. Bland and Altman recommended that 95% of the data should lie within 1.96s of the mean difference [17]. In our data, this is the case, but the limits of agreement (0.87 to -5.71) are wide, reflecting the small sample size and the great variation of the differences.

The difference might be to the following reasons: (1) The research was made with two different designs of the NHPT. the conventional NHPT corresponded to the Smith & Nephew version and the dNHPT corresponds to the version of Mathiowetz [6]. The main difference between these versions is that in one case the container for the pegs consists only of a small round indentation (Smith & Nephew) and in the other case of a deep rectangular container (Mathiowetz). Previous research indicated that there was no difference in the results of the two versions [2], [24], only the study from Mathiowetz et al. reported difficulty picking up pegs from the corners of the square cup [6]. However, the current results indicate that picking up the pegs in the rectangular version was much more difficult for the subjects than the Smith version. The subjects needed more time to pick up the pegs from the large, deep container of the dNHPT. (2) the second factor to consider might have to do with the fact that in the NHPT version the board (blue) has a different color than the pegs (white), so the contrast is higher. The dNHPT shows the same blue color for all components. The pegs are easier to spot in the container and can therefore be picked up more quickly. The pegs of the NHPT are easier to spot in the container and can therefore be picked up more quickly than the pegs of the dNHPT.

B. TEST-RETEST RELIABILITY OF THE DNHPT

The test-retest reliability (ICC (3, k)) of the two dNHPT-sessions (n = 17) was good with healthy adults (ICC: 0.75, [0.28; 0.91]) and significant. The major studies to collect norm data for healthy persons demonstrated only moderate test-retest reliability [2], [6]. However, these studies are not directly comparable because the calculations were made using the Pearson correlation coefficient, which is no longer considered contemporary [11], [25]. In a study comparing healthy people and people with stroke, a significant difference between the ICCs is documented 0.49 and 0.66 in healthy participants and 0.91 and 0.94 in people with stroke [26]. This could indicate that healthy participants achieve a higher variability in the results of the NHPT. However, the large confidence interval could also result from the relatively small sample size.

The reliability level of the ICC from poor to moderate may be related to the insufficient variance of the study subjects. The ICC value indicates what proportion of the total variance over a range of values is due to heterogeneity among study participants [27].

In this study, only healthy subjects of mainly similar age were tested. The lack of variance may result in a lower ICC value.

Furthermore, due to the SDC value found, that changes greater than 2.47 seconds on repeated testing in healthy adults with dNHPT indicates a true change in the probands manual dexterity. Watanabe et al. [26] showed in their study, among others, of health adults comparable results with MDC of 2.2 and 2.6 seconds.

C. INTERRATER RELIABILITY OF THE DNHPT

A high interrater reliability (ICC (2, k)) was obtained for the writing hand (ICC: 0.76) for two measurement points 10 days apart. These results are consistent with previous research reporting good to very good interrater reliability for NHPT. However, all these studies with healthy people calculated the correlation by Pearson (r) or by Spearman (p) and obtained the following results: for healthy adults $r = 0.984$ [2], $r = 0.97$ [6].

Summarizing, our results show high reliability, thus is independent of the tester.

D. CLINICAL IMPLICATIONS

The hand dexterity measurement protocol used in the present work operates strictly according to the standardized specifications after Mathiowetz et al. [6]. Thus, the newly developed prototype enables the standardized procedure, like the conventional NHPT. The dNHPT is not more complicated than the original and does not need additional human and technical requirements. In contrast to the conventional NHPT, the dNHPT has the advantage that users do not need any additional preparation time, as the dNHPT guides them through the entire standardized measurement.

No additional material, such as a stopwatch or documentation material, is required, as these functions are digitally implemented in the dNHPT and shown on a display. This feature allows for a more objective and easier measurement of hand dexterity function. The compact and portable form of the NHPT is also retained in the dNHPT. Therefore, clinicians and also patients can easily use the dNHPT, even in different environments.

The dNHPT provided higher scores (average 2.47 seconds) compared to the original NHPT. At the same time it show good reliability values. Thus, the dNHPT is suitable for its main task, the quantification of changes over time (e.g. to assess a therapy effect). This is because normally the same device is always used for the measurement. Therefore, if the dNHPT is always used, the difference to the original NHPT is not relevant.

E. LIMITATIONS

The study was conducted with healthy individuals without hand dexterity limitations Its findings thus need to be confirmed in future study with patients with hand dexterity. The sample size was calculated to be sufficient for group comparisons according to our power analysis. However,

at the second measurement time point, only n = 17 subjects participated (due to voluntariness of participation), which could affect the strength of the calculations for test-retest reliability and interrater reliability.

Grice et al. [2] suggests that to increase validity, the average of three trials should be used as the valid score of the NHPT. Due to the time limitation of the test protocol and in order to minimize measurement errors due to a possible practice effect, only one measurement with the writing hand was performed with each instrument for all subjects (one measurement with the NHPT and one with the dNHPT at the first measurement time point and one with the dNHPT at the second measurement time point).

In addition, individual outliers were found in the data set. However, since these are not due to implementation errors, but to the natural variability of the subjects, they were left in the data set. It should also be noted that the conventional NHPT has a different color for the pegs than for the container which improves the contrast. This could have an impact on the run time compared to dNHPT. Further studies of dNHPT should test whether changing of colors may affect run time. The next step will be to investigate the practicability of the newly developed measurement instrument. Based on the results of this study, the prototype will be revised or further developed. Afterwards, studies with patients are to be carried out.

V. CONCLUSION

The quality criteria of the dNHPT demonstrated by this study indicate that the dNHPT is suitable for assessing hand dexterity. The additional features of the dNHPT, such as automatic timing and guidance through the standardized measurement, may be beneficial for users and help to further objectify the results of hand dexterity measurement.

VI. CONFLICTS OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGMENT

The authors would like to thank Abdulrahman Khazma for his invaluable input and support throughout the development of the dNHPT prototype and Erna Schönthaler for the support in recruiting and data collection.

In addition, they would like to extend their sincere thanks to all of the participants in their study, who generously shared their time, experiences, and insights with them. Their willingness to engage with their research was essential to the success of this project, and they are deeply grateful for their participation.

REFERENCES

[1] A. L. Fawcett, *Principles of Assessment and Outcome Measurement for Occupational Therapists and Physiotherapists: Theory, Skills and Application*. West Sussex, U.K.: Wiley, 2007.

[2] K. O. Grice, K. A. Vogel, V. Le, A. Mitchell, S. Muniz, and M. A. Vollmer, "Adult norms for a commercially available nine hole peg test for finger dexterity," *Amer. J. Occupational Therapy*, vol. 57, no. 5, pp. 570–573, Sep. 2003, doi: 10.5014/AJOT.57.5.570.

[3] B. Makofske, "Manual dexterity," in *Encyclopedia of Clinical Neuropsychology*. New York, NY, USA: Springer, 2011, doi: 10.1007/978-0-387-79948-3.

[4] Y. A. Smith, "Normative and validation studies of the Nine-Hole Peg Test with children," *Perceptual Motor Skills*, vol. 90, no. 3, pp. 823–843, Jun. 2000, doi: 10.2466/PMS.90.3.823-843.

[5] J. M. I. Zannatha, A. J. M. Tamayo, Á. D. G. Sánchez, J. E. L. Delgado, L. E. R. Cheu, and W. A. S. Arévalo, "Development of a system based on 3D vision, interactive virtual environments, ergonometric signals and a humanoid for stroke rehabilitation," *Comput. Methods Programs Biomed.*, vol. 112, no. 2, pp. 239–249, Nov. 2013, doi: 10.1016/J.CMPB.2013.04.021.

[6] V. Mathiowetz, K. Weber, N. Kashman, and G. Volland, "Adult norms for the Nine Hole Peg Test of finger dexterity," *Occupational Therapy J. Res.*, vol. 5, no. 1, pp. 24–38, Jan. 1985, doi: 10.1177/153944928500500102.

[7] G. M. Johansson and C. K. Häger, "A modified standardized nine hole peg test for valid and reliable kinematic assessment of dexterity post-stroke," *J. Neuroeng. Rehabil.*, vol. 16, no. 1, p. 8, Dec. 2019, doi: 10.1186/S12984-019-0479-Y.

[8] E. G. Xydas and L. S. Louca, "Design and development of a haptic peg-board exercise for the rehabilitation of people with multiple sclerosis," in *Proc. IEEE 10th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2007, pp. 906–916, doi: 10.1109/ICORR.2007.4428532.

[9] O. Lambercy, M.-C. Fluet, I. Lamers, L. Kerkhofs, P. Feys, and R. Gassert, "Assessment of upper limb motor function in patients with multiple sclerosis using the Virtual Peg Insertion Test: A pilot study," in *Proc. IEEE 13th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2013, doi: 10.1109/ICORR.2013.6650494.

[10] B. C. Tobler-Ammann, E. D. de Bruin, M.-C. Fluet, O. Lambercy, R. A. de Bie, and R. H. Knols, "Concurrent validity and test-retest reliability of the virtual peg insertion test to quantify upper limb function in patients with chronic stroke," *J. Neuroeng. Rehabil.*, vol. 13, no. 1, p. 8, Dec. 2016, doi: 10.1186/S12984-016-0116-Y.

[11] H. C. W. De Vet, C. B. Terwee, L. B. Mokkink, and D. L. Knol, *Measurement in Medicine: A Practical Guide*. Cambridge, U.K.: Cambridge Univ. Press, 2011, doi: 10.1017/CBO9780511996214.

[12] C. B. Terwee, C. A. C. Prinsen, M. G. R. Garotti, A. Suman, H. C. W. de Vet, and L. B. Mokkink, "The quality of systematic reviews of health-related outcome measurement instruments," *Qual. Life Res.*, vol. 25, no. 4, pp. 767–779, Sep. 2015, doi: 10.1007/S11136-015-1122-4.

[13] M. Kellor, J. Frost, N. Silberberg, I. Iversen, and R. Cummings, "Hand strength and dexterity," *Amer. J. Occupational Therapy*, vol. 25, no. 2, pp. 77–83, 1971.

[14] P. Feys, I. Lamers, G. Francis, R. Benedict, G. Phillips, N. LaRocca, L. D. Hudson, and R. Rudick, "The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis," *Multiple Sclerosis J.*, vol. 23, no. 5, pp. 711–720, Apr. 2017, doi: 10.1177/1352458517690824.

[15] H.-M. Chen, C. C. Chen, I.-P. Hsueh, S.-L. Huang, and C.-L. Hsieh, "Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke," *Neurorehabil. Neural Repair*, vol. 23, no. 5, pp. 435–440, Jun. 2009, doi: 10.1177/1545968308331146.

[16] S. Plichta and E. Kelvin, *Munro's Statistical Methods for Health Care Research*, 6th ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2013.

[17] J. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307–310, Dec. 1986, doi: 10.1016/S0140-6736(86)90837-8.

[18] D. Giavarina, "Understanding Bland Altman analysis," *Biochem. Med.*, vol. 25, no. 2, pp. 51–141, Jun. 2015, doi: 10.11613/BM.2015.015.

[19] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/J.JCM.2016.02.012.

[20] D. A. van Kampen, W. Willems, L. W. A. H. van Beers, R. M. Castelein, V. A. B. Scholtes, and C. B. Terwee, "Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs)," *J. Orthopaedic Surg. Res.*, vol. 8, no. 1, p. 40, 2013, doi: 10.1186/1749-799X-8-40.

[21] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: Uses in assessing rater reliability," *Psychol. Bull.*, vol. 86, no. 2, pp. 420–428, 1979, doi: 10.1037//0033-2909.86.2.420.
[22] M. C. Fluet, O. Lambercy, and R. Gassert, "Upper limb assessment using a Virtual Peg Insertion Test," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun./Jul. 2011, pp. 1–6, doi: 10.1109/ICORR.2011.5975348.
[23] M. Bowler, F. Amirabdollahian, and K. Dautenhahn, "Using an embedded reality approach to improve test reliability for NHPT tasks," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Jun. 2011, pp. 1–7, doi: 10.1109/ICORR.2011.5975343.
[24] K. Herren and L. Radlinger, "Aussagekraft zweier Versionen des Nine-Hole-Peg-Tests bei Patienten nach Schlaganfall," *Physioscience*, vol. 4, no. 2, pp. 60–64, Jun. 2008, doi: 10.1055/S-2008-1027161.
[25] J. J. Gagnier, J. Lai, L. B. Mokkink, and C. B. Terwee, "COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures," *Qual. Life Res.*, vol. 30, no. 8, pp. 2197–2218, Aug. 2021, doi: 10.1007/s11136-021-02822-4.
[26] N. Watanabe, Y. Otaka, M. Kumagai, K. Kondo, and E. Shimizu, "Reliability of the modified Nine Hole Peg Test in healthy adults and individuals with hemiparetic stroke," *Prog. Rehabil. Med.*, vol. 7, Sep. 2022, Art. no. 20220046, doi: 10.2490/PRM.20220046.
[27] L. G. Portney, *Foundations of Clinical Research: Applications to Evidence-Based Practice*, 4th ed. Philadelphia, PA, USA: Davis Company, 2020.

EVELINE PROCHASKA received the B.Sc. degree in computer engineering and the M.Sc. degree in health assisting engineering from the University of Applied Sciences (UAS) Campus Vienna, Austria, in 2011 and 2014, respectively.

Since 2014, she has been a Lecturer with the Department of Applied Electronics and Technical Informatics, UAS Campus Vienna. She received an endowed professorship in the field of healthcare engineering, in 2021. Her research interests include technology in medicine and medical informatics.

Ms. Prochaska is a member of the Professional Association of Medical Computer Scientists, Germany; the Austrian Association for Electrical Engineering; and the Ethical Commission, UAS Campus Vienna.

ELSKE AMMENWERTH received the Diploma and Ph.D. degrees in medical informatics from the University of Heidelberg, in 1997 and 2001, respectively.

In 2005, she was appointed as a Professor of health informatics with UMIT TIROL–Private University for Health Sciences, Medical Informatics and Technology. She has published over 150 papers and 25 books or proceedings. Her research interests include electronic health records and patient portals, evaluation of IT in health care, evidence-based health informatics, and learning analytics.

Dr. Ammenwerth was a Founder of the Working Group on Assessment of Health Information Systems, European Federation for Health Informatics (EFMI). She is an Austrian Representative of IMIA and EFMI. She is an International Fellow of the American College of Medical Informatics (ACMI) and the Founding Member of the International Academy of Health Science Informatics (IAHSI). She is an Associate Editor of *Applied Clinical Informatics* and a Guest Editor of *Journal of Biomedical Informatics*.