Test-retest reliability and preliminary reference data for the Virtual Eggs Test in healthy adults

Lucia Angelini, Michele Piazzini, Marco Baccini, Tommaso Ciappetti, Francesca Cecchi, Marco Controzzi, *Member*, IEEE

***Abstract*—Rehabilitation professionals rely on hand assessment to select the proper therapy and monitor the recovery process over time. The Virtual Eggs Test (VET) is a unique assessment tool that evaluates both fine and gross hand dexterity, taking into account not only the speed in performing an action but also the accuracy in regulating the grip force. The VET was preliminary validated with amputees and healthy individuals, demonstrating construct validity, but also highlighting some weaknesses that required a protocol revision. Here, we present the update version of the protocol and of the outcomes measure (Fine Dexterity Index – FDI - and Gross Dexterity Index - GDI). We carried on a clinical trial involving 79 healthy participants, stratified by age groups, to collect normative data and evaluate test-retest reliability. The FDI and the GDI demonstrate no significant difference between test and retest. FDI showed excellent reliability (ICC was 0.90 and 0.91 for the dominant and non-dominant hand, respectively). While GDI demonstrated moderate reliability (ICC was equal to 0.63 and 0.64 for the dominant and non-dominant hand, respectively). The findings indicate that the VET is reliable over time for the evaluation of hand function in tasks involving fragile objects.**

***Index Terms*—Hand, Hand evaluation, Motor skills, Occupational Therapy, Rehabilitation**

I. INTRODUCTION

HAND assessment enables rehabilitation professionals to judge the function of the hand [1]. This is crucial for selecting therapy and monitoring progress over time [2][3]. There are several assessment tools that evaluate hand function [4][5], differing in focus and approach, reflecting the complexity of the hand as an organ with a myriad of functions [6]. Hand dexterity is a recognised valid indicator of the function of the hand [7], and it is considered “the best predictor of independence during everyday activities” [8] as a consequence of an impairment. Hand dexterity can be defined as the capability to coordinate voluntary movements (and so forces) to complete a task with precision, speed and efficiency [9]. It can be categorised into fine and gross dexterity depending on the precision of the movement required [10][11]. Hand dexterity evaluation tools have been largely developed and reviewed [2][8][12]. Most of them are time-based tests and often overlook other critical aspects, such as in-hand manipulation or dynamic force control [13]. In order to specifically evaluate dynamic force control, some ad hoc fragile objects or sensorised instruments have been designed alongside their specific protocol [14]-[19]. To the best of our knowledge, none of these tests has been further developed to become a standard hand assessment procedure, except for the Virtual Eggs Test (VET) [18]. The VET is a performance-based assessment tool that examines both fine and gross hand dexterity. The evaluation process considers both the speed of performing the task and the accuracy in regulating the grip force. During the VET, participants are instructed to move objects with varying fragilities from one position to another over a barrier without breaking them and as quickly as possible. The VET has been revised and exploited in different clinical trials involving amputees [18]-[21]. The first validation study of the VET involved amputees and healthy individuals [22], demonstrating construct validity. Test-retest reliability was investigated and demonstrated only for the healthy population [22]. The promising results sprang to enlarge the population of validity to other pathologies (e.g., subjects with stroke outcomes [24]), and to collect normative data for the healthy population, balancing for age and sex [25][26]. Based on feedback, the protocol was revised to improve usability and resolution, while maintaining its 25-minute duration [27]. We preliminary tested the modified version with 7 healthy participants aged 29 to 83 years, with promising results [28]. This study aimed to evaluate the test-retest reliability of the VET in the healthy population and the influence of age and hand-dominance on performance. In addition, it aimed to obtain preliminary normative data for the VET.

This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE.

This work was supported by the Italian Ministry of University and Research, under the complementary actions to the NRRP ‘Fit4MedRob - Fit for Medical Robotics’ Grant (# PNC0000007). *(Corresponding author: M. Controzzi* and *F. Cecchi)*. M.C. and F.C. are equal contribution corresponding authors and L. A. and M. P. are equal contribution first author.

M. Controzzi, L. Angelini are with The BioRobotics Institute, Scuola Superiore Sant’Anna, 56024 Pisa, Italy, and also with the Department of Excellence in Robotics and AI, Scuola Superiore Sant’Anna, 56127 Pisa, Italy (e-mail: lucia.angelini@santannapisa.it and marco.controzzi@santannapisa.it).

M. Piazzini, M. Baccini, T. Ciappetti and F. Cecchi are with IRCCS Fondazione don Carlo Gnocchi ONLUS, Florence, Italy. M. Piazzini and F. Cecchi are also with Dipartimento di Medicina Sperimentale e Clinica, UniFi, Florence, Italy (mail: mpiazzini@dongnocchi.it, marco.baccini@unifi.it, tciapetti@dongnocchi.it and fcecchi@dongnocchi.it)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the ethical committee of Area Toscana Centro, Italy (Approval number: 23327_spe.).

II. METHODS

A. Participants

A convenience sample of healthy subjects was enrolled to collect data from at least 10 participants (5 males and 5 females) for each of the following age groups: 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, ≥80 years. Eligibility criteria were the absence of motor impairments, severe visual and oculomotor deficits or psychiatric comorbidities and a silent history of neurological and musculoskeletal pathologies. The experiment took place at IRCCS (Istituto di Ricerca e Cura a Carattere Scientifico) Fondazione Don Carlo Gnocchi (Florence, Italy), where participants were enrolled on a voluntary basis. The study was conducted following international and national regulations on clinical trials, adhering to the principles of the Declaration of Helsinki. All participants provided informed consent. The study received approval from the local ethical committee of Area Toscana Centro, Italy (Approval number: 23327_spe.).

B. Experimental procedure

Before starting the experiment, the participant's handedness was determined according to the Edinburgh Inventory [30]. Each participant carried out the VET with both hands in two different sessions of assessment (Test and Retest), at a distance of at least 14 days. The side to assess first was randomised among participants and remained constant between sessions. A 3-minute rest pause was allowed after the first hand test.

C. The Virtual Eggs Test

Setup. The setup of the VET consists of 20 Virtual Eggs (VEs), a platform, and software. A VE is a device able to simulate objects with different fragilities: if the applied grip force exceeds the internal break threshold (T_{break}), the external walls collapse and the VE "breaks" reversibly. Each VE is identified by a normalised break threshold defined as the ratio between the T_{break} and the minimum grip force required to lift the object (GF_{min}). GF_{min} is estimated as $W/(2*\mu)$, where W is the weight of the VE and μ is the coefficient of friction between the fingers and the external walls. The external walls are covered with 240-grit sandpaper, yielding an experimentally measured friction coefficient of 1.44. The normalised break thresholds of the 20 VEs range from 1 to 10.5, increasing with a step of 0.5 (TABLE I). We increased the number of VEs comprising the kit from 11 (as in [22]) to 20, and we adjusted their break thresholds to simulate more fragile objects. In addition, we covered the grasp wall with sandpaper to standardize friction and defined the normalize break metric analogously to [23].

The platform hosts three sensorized areas highlighted with a red frame: a central area, *Home position*, and two lateral areas, *right* and *left Egg placers*. The two *Egg placers* are 245mm distant and separated by a 6cm high barrier. During the test, the platform is positioned 10cm from the table edge in front of the participant, who is seated on a height-adjustable chair to keep his/her forearm parallel to the ground when his/her hand is on the platform. The platform connects to a PC where the software runs to visualize and store the data, enabling the examiner to take note of any VEs that are "broken" during the test.

TABLE I
WEIGHTS, BREAK THRESHOLDS AND NORMALISED BREAK THRESHOLDS (Φ) OF THE VES.

VE#	Weight (g)	Break threshold (N)	Φ
1	90.4	0.31	1
2	89.1	0.47	1.5
3	89.2	0.62	2
4	89.6	0.77	2.5
5	90.5	0.92	3
6	90.7	1.09	3.5
7	90.6	1.23	4
8	90.8	1.37	4.5
9	90.9	1.53	5
10	90.4	1.69	5.5
11	90.5	1.86	6
12	90.6	2.02	6.5
13	90.4	2.15	7
14	89.6	2.32	7.5
15	90.9	2.46	8
16	91.0	2.61	8.5
17	90.7	2.77	9
18	90.9	2.94	9.5
19	90.8	3.08	10
20	90.7	3.23	10.5

Protocol. Prior to starting, the participant washes his/her hands. The protocol of the VET includes two consecutive phases, a familiarisation phase and a testing phase, both composed of a number of trials with different VEs. Each trial includes seven consecutive repetitions of VE transport, each consisting of the following sequence of movements: starting with the hand placed on the *Home position*, the subject reaches and grasps the VE from the starting *Egg placer* using a pinch grasp (sub-terminal pinch between thumb and index fingers), moves it over the barrier, places it on the other *Egg placer*, without breaking it at any stage of the trial, and returns the hand over the *Home position*. For the first repetition, the starting *Egg placer* is the one on the side of the tested arm. The ending *Egg placer* of one repetition becomes the starting *Egg placer* in the subsequent repetition. Before each trial, the subject dips the thumb and index finger into magnesite powder and intentionally breaks the VE to test its break threshold (i.e., its fragility).

In the familiarization phase, the participant completes three trials with VE20 (the most robust one), VE10 and VE5, respectively. In the testing phase, the participant performs a variable number of trials with different VEs, starting with VE20. During each trial, a repetition fails if the participant breaks or misplaces the VE, while the trial fails if at least three repetitions fail. After each trial, the VE number for the next trial is determined on the basis of the participant's performance following an up/down method [31], with a step downwards (*after* a successful trial) and upwards (after a failed trial) corresponding to 3 and 2, respectively. If the selected VE number falls below 1 or exceeds 20, VE1 or VE20 is chosen accordingly. The test phase ends either when the fifth reversal (defined as a successful trial after one or more failures, or vice versa) occurs or when the participant has handled VE20 or VE1 for three consecutive trials. This version of the protocol differs from the previous ones because it includes a structured familiarization phase, and an adaptive procedure to select the subsequent VEs. The former was introduced to improve data reliability, whereas the latter was introduced to reduce the duration of the test without compromising the accuracy of the assessment.

Metrics. The VET assesses gross and fine dexterity through two different indices: Fine Dexterity Index (FDI) and Gross Dexterity Index (GDI). Unlike the first validation studies [22][28],where both indices were derived from the parameter of a fitted psychometric curve, in this study, they are computed directly from the the up/down data [31][32] to improve the reliability while preserving the original meaning. Trials are divided based on the first reversal, providing a participant-specific separation between *gross* and *fine* dexterity. Trials before the first reversal contribute to the GDI, while the others to the FDI. This is analogous to the *fragile-robust VEs* division introduced in [22], with improved reliability through participant-specific threshold. For both indices, the lower the value, the better the performance.

GDI is calculated as t_{gross}/p_{gross}, where p_{gross} is the probability of successfully transporting the VEs (defined as the ratio between the number of correct repetitions and the total number of repetitions) and t_{gross} is the minimum flight time among successfully trials. The flight time is defined as the timeframe between lifting the VE from the starting *Egg placer* and releasing it on the destination *Egg placer*. Compared to the previous formulation [22], the revised GDI removes the dependence on a timeout parameter, while maintaining sensitivity both to the movement timeframe and probability of correct transportation.

FDI is obtained by averaging the last four reversals. The normalized threshold (Φ) of the VE corresponding with each reversal is adjusted according to the number of correct repetitions in the trial. Upwards reversals (i.e., a successful trial after one or more failures) and downwards reversals follow equation (1) and equation (2), respectively. The lower limit of the adjusted threshold is 1.

$$\Phi_{\mathrm{VE}} + \frac{s_u * \Delta\Phi}{n_{fa}+1} * n_f = \Phi_{\mathrm{VE}} + \frac{2 * 0.5}{3} * n_f \qquad (1)$$

$$\Phi_{\mathrm{VE}} - \frac{s_d * \Delta\Phi}{n_s - n_{fa}} * (n_s - n_f) = \Phi_{\mathrm{VE}} - \frac{3 * 0.5}{5} * (7 - n_f) \qquad (2)$$

Where s_u and s_d are the steps up and down respectively, n_f the number of failed repetitions, n_{fa} the number of repetitions per session that can be failed without failing the session (i.e., 2), and n_s the number of repetitions per session. In the contest of a high-threshold theory [31] (as in [22] and [28]), the threshold value obtained by the fitting of the psychometric corresponds to the averaging of the last reversals. We revised this value taking into account not only the probability of correct transportation but also the number of failed repetition.

D. Statistical analysis

Statistical analysis was completed using IBM® SPSS® [33] and Matlab [34], to assess test-retest reliability and to evaluate the effect of age and hand-dominance. Descriptive analysis are reported for each group considering all the dataset as median and interquartile range. While, before the other analysis, outliers in each age group were removed to increase the robustness of the analysis itself. For normally distributed data, values outside the interval *mean ± 2 SD,* for non-normally distributed data values outside the interval (Q1 – 1.5*IQR, Q3 + 1.5*IQR), where Q1 and Q3 are the first and the third quartiles, respectively, and IQR is the interquartile range. We tested data distribution for normality using the Kolmogorov-Smirnov test, and homogeneity of variance using the Levene test.

The test-retest reliability of each outcome measure was evaluated separately for the dominant and non-dominant hand. The evaluation was carried out following the CONSMIN guidelines [35][36]. We first evaluated the stability of participants, and then we computed the Intraclass Correlation Coefficient (ICC) with the 95% Confidence Interval [37]. The stability was evaluated with parametric paired t-test or Wilcoxon Signed Rank Test according to the data distribution. We calculated the ICC coefficient as a two-way mixed effects absolute agreement single measurement [37][38]. Based on the criteria of Resnik et al. [12], ICC values were interpreted as follows: excellent reliability for ICC ≥ 0.8, good for 0.80 > ICC ≥ 0.6, and poor if ICC < 0.6. Additionally, we computed the Standard Error Measurement (SEM) (4), Minimal Detectable Change with a 95% CI (MDC_{95}) (5), and MDC_{95} percentage ($MDC_{\%}$) (6), where SD is the pooled standard deviation [39][40] and $\bar{X}$ is the mean of the considered variable. We considered an $MDC_{\%}$ of up to 30% to be acceptable [41].

$$SEM = SD * \sqrt{1 - ICC} \qquad (4)$$

$$MDC_{95} = SEM * 1.96 * \sqrt{2} \qquad (5)$$

$$MDC_{\%} = \frac{MDC_{95}}{\bar{X}} * 100 \qquad (6)$$

To evaluate the effect of age and hand-dominance, we analysed the data of the test phase. When data were normally distributed and homogeneity of variance among groups was met, we performed a mixed Analysis of Variance (ANOVA) (within factor: hand-dominance, between factor: age). We conducted the post-hoc test with Bonferroni correction if one of the factors or their interaction was significant. If data did not meet the requirements of the mixed ANOVA, we performed a two-way non-parametric test following the approach described in [42][43] (within factor: hand dominance, between factor: age). Briefly, it consists of applying a non-parametric test to aggregated data to assess the significance of each factor and its interaction. We selected the Kruskal-Wallis H Test to compare the data of the seven unpaired groups. To compare the data of two paired groups, we choose the T-Wilcoxon Signed Rank if the distribution of the difference between the two populations was symmetric (|γ| < 1, where γ is the skewness index), otherwise, we used the Sign test. When a factor or its interaction was found to be significant, post-hoc non-parametric tests with

Bonferroni correction were performed. The statistical significance of all the tests performed was considered by $p<0.05$.

III. RESULTS

Eighty-three participants were enrolled in the study; 4 participants dropped out and did not complete the protocol, so a total of 79 participants were considered for the analysis. Participants were categorized into seven age-based groups: 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, ≥80 (TABLE II). The first participant for each age-based group was previously analysed in a pilot study aimed at preliminary evaluating the protocol

TABLE II

PARTICIPANT ENROLLED IN THE STUDY.

Group age	Average age	Number of Male	Number of Female	Total Number	Number of right-handed
18–29	26.2	5	7	12	10
30–39	31.4	5	5	10	9
40–49	46.5	5	5	10	8
50–59	55.6	5	6	11	11
60–69	63.8	8	5	13	12
70–79	74.7	6	7	13	12
≥80	82.9	5	5	10	9

TABLE III

MEDIAN AND INTERQUARTILE RANGE OF THE OUTCOME MEASURE OF THE VET FOR EACH AGE GROUP, HAND AND SESSIONS (I.E., TEST AND RETEST). VARIABLES THAT ARE NOT NORMALLY DISTRIBUTED ARE REPORTED IN BOLD.

	FDI				GDI			
	Dominant Hand		Non-dominant hand		Dominant Hand		Non-dominant hand	
Group age	Test	Retest	Test	Retest	Test	Retest	Test	Retest
18–29	**1.54±0.12**	**1.50±0.12**	1.62±0.12	1.60±0.12	0.63±0.22	0.65±0.14	0.69±0.26	0.68±0.17
30–39	1.58±0.21	1.50±0.21	1.60±0.21	**1.58±0.17**	0.68±0.17	0.61±0.11	0.69±0.14	0.61±0.09
40–49	1.60±0.29	1.54±0.21	**1.70±0.12**	**1.56±0.12**	0.70±0.16	0.75±0.14	0.76±0.19	0.73±0.13
50–59	1.58±0.14	1.58±0.27	1.58±0.10	**1.50±0.16**	0.71±0.13	0.71±0.15	0.72±0.28	0.77±0.16
60–69	1.79±0.24	1.71±0.22	1.79±0.42	**1.71±0.39**	0.74±0.24	0.73±0.18	0.72±0.17	0.70±0.18
70–79	2.25±0.67	2.17±0.98	2.35±0.87	2.25±1.00	0.76±0.11	0.74±0.12	0.78±0.20	0.78±0.12
≥80	2.54±1.09	2.47±1.17	2.50±0.62	2.54±1.17	0.84±0.11	0.85±0.10	0.85±0.10	0.87±0.16

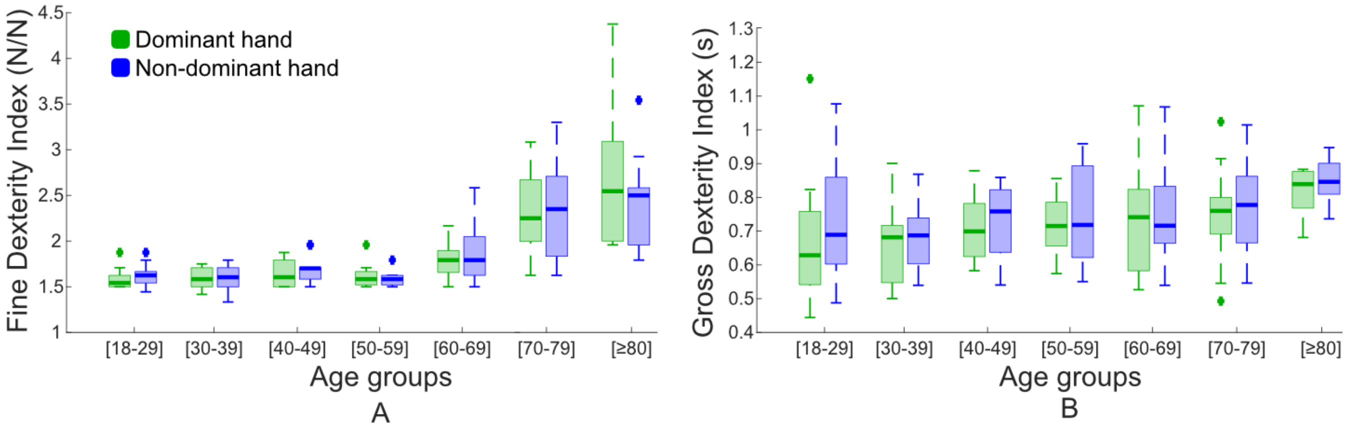

Fig. 1 Outcome measures of the Virtual Eggs Test (VET) for the test session and different age groups. Fine Dexterity (FDI) (A), Gross Dexterity Index (GDI) (B). Green circles represent average value of the dominant hand, blue rhombuses represent average value of the non-dominant hand.

TABLE IV

Test-retest reliability for Fine Dexterity Index and Gross Dexterity Index. Test and Retest mean and SD, number of subject per group (N) after removing outliers, results of Wilcoxon Signed Rank Test (Z) and results of Paired t-test (t) and Coefficient Interval (CI)[29], Intraclass Correlation Coefficient (ICC) and 95% confidence interval (95% CI), Standard Error Measurement (SEM), Minimal Detectable Change with a 95% CI (MDC_{95}), and MDC_{95} percentage ($MDC_{\%}$).

	FDI		**GDI**	
	Dominant hand	**Non-Dominant hand**	**Dominant hand**	**Non-Dominant hand**
Test	1.90 ± 0.56	1.86 ± 0.44	0.71 ± 0.12	0.74 ± 0.13
Retest	1.86 ± 0.54	1.86 ± 0.51	0.70 ± 0.12	0.73 ± 0.12
N	74	72	73	75
Statistic	Z = 1.571 p = 0.116 CI = [0; 0.05]	Z = 0.418 p = 0.676 CI = [-0.05; 0.05]	t = 0.301 p = 0.759 CI = [-0.02; 0.03]	t = 1.081 p = 0.283 CI = [-0.01; 0.04]
ICC (95% CI)	0.90 (0.84-0.93)	0.91 (0.85-0.94)	0.63 (0.47-0.65)	0.64 (0.48-0.65)
SEM	0.17	0.14	0.07	0.08
MDC_{95}	0.48	0.40	0.21	0.21
$MDC_{\%}$	25.7	21.8	29.5	29.0

reliability [28]. About 80% of participants were right-handed according to the Edinburgh Inventory.

FDI and GDI are reported according to the hand-dominance, the age and the session (i.e., Test or Retest) in TABLE III and Fig. 1. The mean distance between test and retest was 21.5±8.9 days.

Data on test-retest reliability, evaluated for each index and each hand separately, are shown in TABLE IV. For all variables, no significant differences were detected between test and retest. ICCs were similar for the two hands and were ≥0.90 for FDI and ≥0.60 for GDI. The MDC% was always below 30%.

Focusing on the data of the test session, we evaluated the influence of age and hand-dominance. With regard to FDI, the factor hand-dominance had no significant effect (Sign test, p-value = 0.504), as the interaction between age and hand (Kruskal-Wallis H Test, p-value = 0.359). Conversely, FDI was significantly influenced by the factor age (Kruskal-Wallis H Test, p-value < 0.001), with the oldest group (age ≥80) being different from groups 18 to 69, the group [70-79] being different from the groups 18 to 59 (TABLE V). Concerning the GDI, a significant effect was found for both the hand-dominance factor (ANOVA, p-value < 0.001), with values significantly lower in the dominant than in the non-dominant hand, and the age factor (ANOVA, p-value = 0.035). For the latter, the post hoc analysis showed that only the difference between the oldest and the youngest group was significant (p-value adjusted with Bonferroni = 0.037). No significant differences were found considering the interaction between hand and age.

IV. Discussion

The VET is a performance-based evaluation test that assesses both gross and fine dexterity, focusing on the ability of the participant to transport fragile objects without crushing them. The VET was validated for the first time with 30 myoelectric hand users [22]. The validation highlighted both the novelty of the assessment tool and its limitations, underscoring the need for modifications to the hardware and procedure. Here, we report preliminary normative data for the healthy population of 79 adults acquired considering the new protocol of the VET and the new definition of FDI and GDI. Both indices show good to excellent test-retest reliability [12]. FDI has an ICC of 0.90 and 0.91, respectively, for the dominant and non-dominant hand, indicating excellent reliability. Those ICC values also meet the requirement of Fitzpatrick et al. [44], allowing the score of the test to be used as a basis for individual decision. GDI shows a good reliability for both hands (ICC = 0.63 and 0.64 for the dominant and non-dominant hand, respectively). Even if the Fitzpatrick's threshold of 0.90 [44] is not reached, those ICC values are comparable with some assessments commonly employed in clinical practice [11][45][46]. In this context, the results obtained can be considered acceptable. Nevertheless, additional clinical trials should be carried out to assess these properties. Increasing the number of participants for each age group, reducing the variability in the interval between test and retest may improve the reliability of outcomes. The difference between the reliability of FDI and GDI can suggest that the formulation of GDI could be further improved and/or that a ceiling effect may be present. GDI derived from the ratio of t_{gross}/p_{gross}, and it is likely to assume that the probability of correct transportation (p_{gross}) is always close to 1 in the healthy participants.

No statistically significant differences are highlighted for FDI between the dominant and non-dominant hands independently of the age group (since no significance has been found for the hand or the hand*age factor). Since FDI is more related to fragile objects, this finding is consistent with what was assessed by Gorniak et al. during the study on the manipulation of fragile objects [23]. On the contrary, GDI values indicate on average better performance of the dominant hand. This is consistent with the results of other manual dexterity tests focusing on movement speed, in which the dominant hand usually performs significantly better than the non-dominant hand. This has been demonstrated, for example, for the Box and Block test [47][48], the Functional Dexterity Test [49] and the Nine Hole Peg Test

TABLE V

Statistical results for Fine Dexterity Index and Gross Dexterity Index. Results of ANOVA tests (F) and its post-hoc comparisons, results of Sign test (Z), and results of the Kruskal-Wallis H Test (χ^2) and its post-hoc comparisons with Mann-Whitney U test. For the post-hoc test, p-values (p_{adj}) are adjusted with the Bonferroni correction. Significant results (p<0.05) are highlighted in boldface. Number of subject per group (n) after removing outliers.

			FDI	GDI
		N	73	76
Main Effects		Hand	$Z = -0.668$ $p = 0.504$ CI = [0.04; 0,00] $ES = 0.078$	**F = 14.218** **$p < 0.001$** **CI = [0.701;0,753]** **$ES = 0.171$**
		Age	**$\chi^2 = 45.393$** **$p < 0.001$** **$ES = 0.597$**	**F = 2.417** **$p = 0.035$** **$ES = 0.174$**
		Hand*age	$\chi^2 = 6.603$ $p = 0.359$ ES = 0.009	F = 0.564 $p = 0.757$ $ES = 0.047$
Post-Hoc	Age	[18– 29] – [30 – 39]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[18– 29] – [40 – 49]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[18– 29] – [50 – 59]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[18– 29] – [60 – 69]	$p_{adj} = 0.170$	$p_{adj} = 1.000$
		[18– 29] – [70 – 79]	**$p_{adj} = 0.002$**	$p_{adj} = 0.756$
		[18–29] – [≥80]	**$p_{adj} = 0.005$**	**$p_{adj} = 0.037$**
		[30–39] – [40–49]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[30–39] – [50–59]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[30–39] – [60–69]	$p_{adj} = 0.400$	$p_{adj} = 1.000$
		[30–39] – [70–79]	**$p_{adj} = 0.004$**	$p_{adj} = 1.000$
		[30–39] – [≥80]	**$p_{adj} = 0.005$**	$p_{adj} = 0.085$
		[40–49] – [50–59]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[40–49] – [60–69]	$p_{adj} = 1.000$	$p_{adj} = 1.000$
		[40–49] – [70–79]	**$p_{adj} = 0.013$**	$p_{adj} = 1.000$
		[40–49] – [≥80]	**$p_{adj} = 0.007$**	$p_{adj} = 0.781$
		[50–59] – [60–69]	$p_{adj} = 0.262$	$p_{adj} = 1.000$
		[50–59] – [70–79]	**$p_{adj} = 0.013$**	$p_{adj} = 1.000$
		[50–59] – [≥80]	**$p_{adj} = 0.007$**	$p_{adj} = 1.000$
		[60–69] – [70–79]	$p_{adj} = 0.098$	$p_{adj} = 1.000$
		[60–69] – [≥80]	**$p_{adj} = 0.015$**	$p_{adj} = 0.489$
		[70–79] – [≥80]	$p_{adj} = 1.000$	$p_{adj} = 1.000$

[50]. Likely, this difference is related to the greater strength and coordination of the dominant limb, which allows for faster and more forceful movements, a key element in completing the transport of objects in the shortest possible time.

FDI shows statistical differences in terms of age, with older people showing less dexterous than young adults. This trend was also found for GDI, even if statistical results indicate differences only between the group of ≥80 and the group of 18-29. The trend is coherent with normative data collected for other dexterity tests as the Nine Hole Peg Test [51] or the Box and Block Test [10], as well as with data from specific studies [25][26] that investigate factors influencing dexterity. For GDI, both SEM and MDC_{95} are comparable between the two hands, indicating that the VET can assess the dominant and non-dominant hands with similar precision. Those values can in addition be used to evaluate the minimal not-aleatory change in two different administrations of the VET [52]. The $MDC_{\%}$ of both FDI and GDI fall within acceptable limits for both hands [41]. The MDC% allows clinicians to distinguish clinical relevant changes from random variation. In detail, if the difference in terms of percentage between two repeated measurements exceeds the MDC%, then the change is unlikely to be due to chance and can be considered clinical relevant.

The findings prove the test-retest reliability of the VET in the healthy population. This candidates the VET as the first test capable of assessing hand dexterity in tasks requiring the manipulation of fragile objects. Further studies are needed to complete the validation of the VET. Firstly, expanding the number of people will increase the statistical power of the study, and allow for the investigation of gender-related differences and the establishment of robust normative data. The sample size recruited for this study can be considered acceptable for the main scope of the study according to COSMIN recommendations, which indicate a sample size of at least >50 participants to be considered *adequate* [53]. However, it is certainly insufficient to fully explore the effect of age, gender and other variables, such as work activity, on performance. Therefore, the number of participants is the main limitation of the present work. Other important properties such as validity and responsiveness should be evaluated. Verifying the validity of the VET is of paramount importance, since an inherent limitation of this test is the long time required for its administration (about 20 minutes). This administration time is much higher that the time required to complete classic dexterity tests such as the Box & Block Test [54]. In our opinion, however, this could certainly be acceptable if the test proves to be not only reliable but also capable of providing additional insights into the person's manual dexterity that cannot be detected by traditional tests.

V. CONCLUSIONS

Among the other hand assessment tools, the VET emerges for its peculiarity of assessing the hand function by evaluating the transport of fragile objects. The VET evaluates hand dexterity based on the accuracy of the task and on the capability of regulating the grip force. In this study, we presented the new protocol of the VET alongside the new definition of the FDI and the GDI, that showed good to excellent reliability in the healthy population. Preliminary normative data were also collected in different age groups. Additional studies are needed to confirm these findings in a larger sample of healthy participants, to explore other metric properties of the VET and to validate it in other populations.

ACKNOWLEDGMENT

L.A. and M.C. gratefully acknowledge Robinson Guachi Guachi for his contribution to the development of the Virtual Eggs.

DATA AVAILABILITY

All data are available at zenodo.18553401.

REFERENCES

[1] J. Bear-Lehman and B. C. Abreu, “Evaluating the Hand: Issues in Reliability and Validity,” *Physical Therapy*, vol. 69, no. 12, pp. 1025–1033, Dec. 1989, doi: https://doi.org/10.1093/ptj/69.12.1025.

[2] K. Schoneveld, H. Wittink, and T. Takken, “Clinimetric Evaluation of Measurement Tools Used in Hand Therapy to Assess Activity and Participation,” *Journal of Hand Therapy*, vol. 22, no. 3, pp. 221–236, Jul. 2009, doi: https://doi.org/10.1016/j.jht.2008.11.005.

[3] L. B. Mokkink *et al.*, “The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study,” *Quality of Life Research*, vol. 19, no. 4, pp. 539–549, Feb. 2010, doi: https://doi.org/10.1007/s11136-010-9606-8.

[4] L. A. Jones, “The assessment of hand function: A critical review of techniques,” *The Journal of Hand Surgery*, vol. 14, no. 2, pp. 221–228, Mar. 1989, doi: https://doi.org/10.1016/0363-5023(89)90010-5.

[5] B. M. Haward and M. J. Griffin, “Repeatability of grip strength and dexterity tests and the effects of age and gender,” *International Archives of Occupational and Environmental Health*, vol. 75, no. 1, pp. 111–119, Jan. 2002, doi: https://doi.org/10.1007/s004200100285.

[6] A. R. Sobinov and S. J. Bensmaia, “The neural mechanisms of manual dexterity,” *Nature Reviews Neuroscience*, vol. 22, no. 12, pp. 1–17, Oct. 2021, doi: https://doi.org/10.1038/s41583-021-00528-7.

[7] F. Poirier, “Dexterity as a Valid Measure of Hand Function,” *Occupational Therapy In Health Care*, vol. 4, no. 3–4, pp. 69–83, Jan. 1988, doi: https://doi.org/10.1080/j003v04n03_07.

[8] K. E. Yancosek and D. Howell, “A Narrative Review of Dexterity Assessments,” *Journal of Hand Therapy*, vol. 22, no. 3, pp. 258–270, Jul. 2009, doi: https://doi.org/10.1016/j.jht.2008.11.004.

[9] J. Yong, J. C. MacDermid, and T. Packham, “Defining dexterity—Untangling the discourse in clinical practice,” *Journal of Hand Therapy*, vol. 33, no. 4, pp. 517–519, Oct. 2020, doi: https://doi.org/10.1016/j.jht.2019.11.001.

[10] J. Desrosiers, G. Bravo, R. Hébert, É. Dutil, and L. Mercier, “Validation of the Box and Block Test as a measure of dexterity of elderly people: Reliability, validity, and norms studies,” *Archives of Physical Medicine and Rehabilitation*, vol. 75, no. 7, pp. 751–755, Jul. 1994, doi: https://doi.org/10.1016/0003-9993(94)90130-9.

[11] J. Desrosiers, R. Hébert, G. Bravo, and E. Dutil, “The Purdue Pegboard Test: Normative data for people aged 60 and over,” *Disability and Rehabilitation*, vol. 17, no. 5, pp. 217–224, Jan. 1995, doi: https://doi.org/10.3109/09638289509166638.

[12] L. Resnik, M. Borgia, B. Silver, and J. Cancio, “Systematic Review of Measures of Impairment and Activity Limitation for Persons With Upper Limb Trauma and Amputation,” *Archives of Physical Medicine and Rehabilitation*, vol. 98, no. 9, pp. 1863-1892.e14, Sep. 2017, doi: https://doi.org/10.1016/j.apmr.2017.01.015.

[13] S. V. Duff, D. H. Aaron, G. R. Gogola, and F. J. Valero-Cuevas, “Innovative evaluation of dexterity in pediatrics,” *Journal of Hand Therapy*, vol. 28, no. 2, pp. 144–150, Apr. 2015, doi: https://doi.org/10.1016/j.jht.2015.01.004.

[14] S. G. Meek, S. C. Jacobsen, and P. P. Goulding, “Extended physiologic taction: design and evaluation of a proportional force feedback system.,” *PubMed*, vol. 26, no. 3, pp. 53–62, Jan. 1989.

[15] T. C. Hansen, M. A. Trout, J. L. Segil, D. J. Warren, and J. A. George, “A Bionic Hand for Semi-Autonomous Fragile Object Manipulation via Proximity and Pressure Sensors,” *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Nov. 2021, doi: https://doi.org/10.1109/embc46164.2021.9629622.

[16] M. Sartori, G. Durandau, S. Došen, and D. Farina, “Robust simultaneous myoelectric control of multiple degrees of freedom in wrist-hand prostheses by real-time neuromusculoskeletal modeling,” *Journal of Neural Engineering*, vol. 15, no. 6, p. 066026, Oct. 2018, doi: https://doi.org/10.1088/1741-2552/aae26b.

[17] P. G. Sagastegui, A. Boesendorfer, O. C. Aszmann, J. Ibáñez, and D. Farina, “Excitation of natural spinal reflex loops in the sensory-motor control of hand prostheses,” *Science Robotics*, vol. 9, no. 90, May 2024, doi: https://doi.org/10.1126/scirobotics.adl0085.

[18] F. Clemente, M. D’Alonzo, M. Controzzi, B. B. Edin, and C. Cipriani, “Non-Invasive, Temporally Discrete Feedback of Object Contact and Release Improves Grasp Control of Closed-Loop Myoelectric Transradial Prostheses,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 12, pp. 1314–1322, Dec. 2016, doi: https://doi.org/10.1109/tnsre.2015.2500586.

[19] J. Hermsdörfer, E. Hagl, D. A. Nowak, and C. Marquardt, “Grip force control during object manipulation in cerebral stroke,” *Clinical Neurophysiology*, vol. 114, no. 5, pp. 915–929, May 2003, doi: https://doi.org/10.1016/s1388-2457(03)00042-7.

[20] E. Mastinu *et al.*, “Grip control and motor coordination with implanted and surface electrodes while grasping with an osseointegrated prosthetic hand,” *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, Apr. 2019, doi: https://doi.org/10.1186/s12984-019-0511-2.

[21]F. M. Petrini *et al.*, “Six-Month Assessment of a Hand Prosthesis with Intraneural Tactile Feedback,” *Annals of Neurology*, vol. 85, no. 1, pp. 137–154, Dec. 2018, doi: https://doi.org/10.1002/ana.25384.

[22]M. Controzzi *et al.*, “Assessing Hand Function in Trans-Radial Amputees Wearing Myoelectric Hands: The Virtual Eggs Test (VET),” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 286–297, Dec. 2024, doi: https://doi.org/10.1109/tnsre.2024.3524791.

[23]S. L. Gorniak, V. M. Zatsiorsky, and M. L. Latash, “Manipulation of a fragile object,” *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, vol. 202, no. 2, pp. 413–430, Apr. 2010, doi: https://doi.org/10.1007/s00221-009-2148-z.

[24]M. Piazzini *et al.*, “Application and adaptation of the Virtual Eggs Test for assessing hand dexterity in subjects with stroke: a pilot study,” *Archives of Physiotherapy*, pp. 131–137, Jun. 2025, doi: https://doi.org/10.33393/aop.2025.3291.

[25]Yogi Tri Prasetyo, “Factors Affecting Gross Manual Dexterity: A Structural Equation Modeling Approach,” Apr. 2020, doi: https://doi.org/10.1109/iciea49774.2020.9102096.

[26]Y. T. Prasetyo, “Factors Affecting Fine Dexterity,” *Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering*, Apr. 2020, doi: https://doi.org/10.1145/3396743.3396776.

[27]L. Angelini, Robinson Guachi, and M. Controzzi, “Improvements on the Virtual Eggs Test to Assess Hand Dexterity,” *Biosystems & biorobotics*, pp. 177–180, Jan. 2025, doi: https://doi.org/10.1007/978-3-031-77588-8_34.

[28]L. Angelini *et al.*, “Enhancing Hand Dexterity Assessment Through Protocol Revision and Pilot Evaluation of the Virtual Eggs Test,” *PubMed*, vol. 2025, pp. 205–210, May 2025, doi: https://doi.org/10.1109/icorr66766.2025.11063032.

[29]M. Tomczak and E. Tomczak. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends in Sport Sciences. 2014; 1(21):19-25

[30]R. C. Oldfield, “The assessment and analysis of handedness: The Edinburgh inventory,” *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, Mar. 1971, doi: https://doi.org/10.1016/0028-3932(71)90067-4.

[31]Nicolaas Prins and F. A. A. Kingdom, *Psychophysics A Practical Introduction*. San Diego Elsevier Science & Technology Ann Arbor, Michigan Proquest, 2016.

[32]B. Treutwein and H. Strasburger, “Fitting the psychometric function,” *Perception & Psychophysics*, vol. 61, no. 1, pp. 87–106, Jan. 1999, doi: https://doi.org/10.3758/bf03211951

[33]IBM SPSS Statistics for Windows. Version 20.0, IBM Corp., 2011.

[34]MATLAB. Version 24.2.0 (R2024b) Update 4, The MathWorks, Inc., 2024.

[35]L. B. Mokkink *et al.*, “COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study,” *BMC Medical Research Methodology*, vol. 20, no. 1, Dec. 2020, doi: https://doi.org/10.1186/s12874-020-01179-5.

[36]J. E. Lexell and D. Y. Downham, “How to Assess the Reliability of Measurements in Rehabilitation,” *American Journal of Physical Medicine & Rehabilitation*, vol. 84, no. 9, pp. 719–723, Sep. 2005, doi: https://doi.org/10.1097/01.phm.0000176452.17771.20.

[37]K. O. McGraw e S. P. Wong, “Forming inferences about some intraclass correlation coefficients.”, *Psycholog. Methods*, vol. 1, n. 1, pp. 30–46, 1996. : https://doi.org/10.1037/1082-989x.1.1.30

[38]T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *Journal of Chiropractic Medicine*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: https://doi.org/10.1016/j.jcm.2016.02.012.

[39]W. de, C. B. Terwee, L. B. Mokkink, and D. L. Knol, *Measurement in Medicine*. Cambridge University Press, 2011.

[40]“Foundations of Clinical Research: Applications to Practice,” *Survey of Ophthalmology*, vol. 47, no. 6, p. 598, Nov. 2002, doi: https://doi.org/10.1016/s0039-6257(02)00362-4.

[41]H.-M. Chen, C. C. Chen, I-Ping. Hsueh, S.-L. Huang, and C.-L. Hsieh, “Test-retest reproducibility and smallest real difference of 5 hand function tests in patients with stroke,” *Neurorehabilitation and neural repair*, vol. 23, no. 5, pp. 435–40, 2009, doi: https://doi.org/10.1177/1545968308331146.

[42]G. G. Koch, “The Use of Non-Parametric Methods in the Statistical Analysis of a Complex Split Plot Experiment,” *Biometrics*, vol. 26, no. 1, p. 105, Mar. 1970, doi: https://doi.org/10.2307/2529048.

[43]G. Landenna, and D. Marasini. Metodi statistici non parametrici. il Mulino, 1990, 511 p.

[44]Fitzpatrick, Davey, Buxton, and Jones, “Evaluating patient-based outcome measures for use in clinical trials.,” *Health Technology Assessment*, vol. 2, no. 14, 1998, doi: https://doi.org/10.3310/hta2140.

[45]A. Michimata, T. Kondo, Y. Suzukamo, M. Chiba, and S.-I. Izumi, “The Manual Function Test: Norms for 20- to 90-Year-Olds and Effects of Age, Gender, and Hand Dominance on Dexterity,” *The Tohoku Journal of Experimental Medicine*, vol. 214, no. 3, pp. 257–267, 2008, doi: https://doi.org/10.1620/tjem.214.257.

[46]L. Tesio, A. Simone, G. Zebellin, V. Rota, C. Malfitano, and L. Perucca, “Bimanual dexterity assessment,” *International Journal of Rehabilitation Research*, vol. 39, no. 1, pp. 57–62, Mar. 2016, doi: https://doi.org/10.1097/mrr.0000000000000145.

[47]K.-Y. Li, L.-J. Lin, A.-T. Chan, C.-H. Chen, W.-M. Chang, and Y.-J. Cho, “Population based norms for the box and blocks test in healthy right-handed Taiwanese adults,” *Biomedical Journal*, vol. 43, no. 6, pp. 484–489, Dec. 2020, doi: https://doi.org/10.1016/j.bj.2019.10.004.

[48]D. Natta, E. Alagnidé, T. Kpadonou, C. Detrembleur, T. Lejeune, and G. Stoquart, “Box and block test in Beninese adults,” *Journal of Rehabilitation Medicine*, vol. 47, no. 10, pp. 970–973, 2015, doi: https://doi.org/10.2340/16501977-2023.

[49]F. Şahin, N. Ş. Atalay, N. Akkaya, and S. Aksoy, “Factors affecting the results of the functional dexterity test,” *Journal of Hand Therapy*, vol. 30, no. 1, pp. 74–79, Jan. 2017, doi: https://doi.org/10.1016/j.jht.2016.04.005.

[50]R. Moorthy, S. John, T. Sebastian, and V. Rajshekhar, “Evaluation of hand function in healthy individuals and patients undergoing uninstrumented central corpectomy for cervical spondylotic myelopathy using nine-hole peg test,” *Neurology India*, vol. 65, no. 5, p. 1025, 2017, doi: https://doi.org/10.4103/neuroindia.ni_12_17.

[51]V. Mathiowetz, K. Weber, N. Kashman, and G. Volland, “Adult Norms for the Nine Hole Peg Test of Finger Dexterity,” *The Occupational Therapy Journal of Research*, vol. 5, no. 1, pp. 24–38, Jan. 1985, doi: https://doi.org/10.1177/153944928500500102.

[52]L. Furlan e A. Sterr, “The Applicability of Standard Error of Measurement and Minimal Detectable Change to Motor Learning Research—A Behavioral Study”, *Frontiers Human Neurosci.*, vol. 12, mar. 2018. [Online]. Disponibile: https://doi.org/10.3389/fnhum.2018.00095S. S. Rane, A. Paul, and R. P. Mullerpatan, “Reference values of manual dexterity using Minnesota Rate of Manipulation Test for Indian adults,” *Hand Therapy*, vol. 22, no. 2, pp. 79–85, Feb. 2017, doi: https://doi.org/10.1177/1758998317694474.

[53]C. A. C. Prinsen *et al.*, “COSMIN guideline for systematic reviews of patient-reported outcome measures”, *Qual. Life Res.*, vol. 27, n. 5, pp. 1147–1157, feb. 2018. [Online]. Disponibile: https://doi.org/10.1007/s11136-018-1798-3

[54]V. Mathiowetz, G. Volland, N. Kashman e K. Weber, “Adult Norms for the Box and Block Test of Manual Dexterity”, *Amer. J. Occupational Therapy*, vol. 39, n. 6, pp. 386–391, giu. 1985. [Online]. Disponibile: https://doi.org/10.5014/ajot.39.6.386