Graphical tasks to aid in diagnosing and monitoring Parkinson's disease

Smits, Esther

CHAPTER 4

REPRODUCIBILITY OF STANDARDIZED FINE MOTOR CONTROL TASKS AND AGE EFFECTS IN HEALTHY ADULTS

Esther J. Smits
Antti J. Tolonen
Luc Cluitmans
Mark van Gils
Rutger C. Zietsma
Marina A.J. Tijssen
Natasha M. Maurits

Under review at the journal Measurement (2015)

4.1. Abstract

Standardized handwriting and drawing tasks provide objective measures of important motor symptoms of Parkinson's disease (PD) and allow distinguishing PD patients from age-matched healthy controls. Such tasks could potentially be useful in clinical settings for (early) diagnosis and monitoring purposes. The aim of this study was to assess the reproducibility of a set of standardized graphical tasks including age effects in healthy adults (20–75 years). Overall, movement time and accuracy on circle, spiral and zigzag drawing tasks and a modified Fitts' task showed good reproducibility (intraclass correlation coefficients>0.7). Reproducibility was similar to the reproducibility of the Purdue pegboard task, which is an already validated fine motor control task. Reproducibility was higher in older participants (56–75 years) compared to younger participants (20–55 years). To conclude, performance on this set of standardized graphical tasks was reproducible in healthy adults, which is essential for future diagnostic and monitoring use in patients.

4.2. Introduction

Despite the increased use of computers, the use of a pen for handwriting and drawing is still an important skill in daily life that everyone is expected to master. Holding a pen and performing handwriting and drawing is one of the most complex fine motor functions of humans[1], involving a cooperation between the central nervous system (CNS) and the musculoskeletal system[2]. Therefore, deficits in brain function or in the musculoskeletal system due to a disease, such as Parkinson's disease (PD)[3] or trauma could cause deterioration in handwriting and drawing ability[2]. Typically, handwriting and drawing becomes smaller (bradykinesia) and slower (micrographia) due to PD[3–8]. Even though handwriting and drawing are complex functions, these graphical tasks entail overlearnt skills[9]. Therefore, once mastered, performance on such tasks is expected to not considerably improve or deteriorate over time anymore[9]. Because of this expected stability in performance, graphical tasks are interesting to study to gain more insight into the changes in motor control due to a movement disorder, such as PD[4,6,10], to evaluate treatment effects[11,12] or to study fine motor control in general[1,13–17]. Additionally, graphical tasks have been investigated in several studies as an aid in the diagnostic process of PD, to distinguish PD from healthy controls (HC)[5,7,18,19], or PD from other movement disorders[8]. Before such tasks can be used clinically for diagnosis or screening, their characteristics and added value should be assessed[20]. According to Van den Bruel et al.[20], several steps should be followed in this process. One of these steps is to examine the reproducibility of the results, defined as the ability to achieve the same test results on repeated testing[20]. Only one of the previously mentioned studies examined reproducibility of handwriting tasks[17]. Two other studies investigated the reproducibility of drawing and tapping tasks employing a graphic tablet[21,22]. However, the scope of these previous studies was limited. Mergl et al.[17] investigated reproducibility in young adults (n=21) only and their measures focused on movement speed. Erasmus et al.[21] only investigated reproducibility of drawing precision between two consecutive days. Finally, Feys et al.[22] focused on tremor measures and only investigated short-term test-retest reliability for a spiral drawing task in MS patients with tremor. Therefore, the goal of the present study was to investigate the reproducibility of a set of standardized graphical tasks, executed with the same pen and tablet, with a one-week interval in healthy participants of different ages. Additionally, in the present study intraclass correlation coefficients (ICC) were used to determine the reproducibility, instead of Pearson or Spearman correlations which were used in the earlier studies. For the ICC, the data are centred and scaled using a pooled mean and standard deviation, whereas for the Pearson or Spearman correlation coefficient, each variable is centred and scaled by its own mean and standard deviation. Because measurements on repeated testing are of the same quantity and unit, the ICC is a better measure to examine reproducibility than the Pearson or Spearman correlation coefficient. Furthermore, since some movement disorders are typically diagnosed in specific age-groups – e.g., PD is typically diagnosed in persons older than 60 years – in this study also the influence of age on reproducibility of these tasks was examined.

In a previous study[5] we investigated whether a specific set of tracing tasks could be used to distinguish PD patients from age-matched HC participants. Performance on these tracing tasks was not significantly different between PD patients and HC participants, however, this might be due to the very strict Bonferroni correction which was used to correct for multiple comparisons. Uncorrected, the movement time on the simple tracing tasks did show a significant difference between PD patients and HC participants ($p<0.01$ for all tasks). Therefore the reproducibility of some of these tracing tasks and additional fine motor control tasks using a digital pen and tablet was investigated for the present study in healthy adults of different ages. The additional fine motor control tasks were added to the task-battery of our previous study, which only included simple circle, line and spiral tracing tasks, to cover a larger range of upper limb functions. The line drawing task was replaced by a zigzag tracing and drawing task, because for the zigzag tasks movements must be made in different directions, which require more complex movements of the fingers, wrist and arm[4] than the line tracing task. A modified Fitts' task was also added to the set of tasks. Fitts' task is an extensively used task to study fine motor control of the upper limb. Fitts[23] and others[24,25] reported that when healthy participants performed rapidly alternating movements of the upper limb, movement time varied systematically with changes in movement amplitude and target width (accuracy was held constant). This relationship is currently known as 'Fitts' law'[23]. Changes in this so-called speed-accuracy trade-off could reflect changes in fine motor control of the upper limb and therefore Fitts' task offers a useful framework in which movement processes can be studied, both in general and at clinical level[26]. For example, Sanes[27] and Rand et al.[28] showed that movement time and inaccuracy of PD patients were larger compared to HC participants in a modified Fitts' task. Here, we explored which of these tasks showed the best reproducibility.

To show that the new task-battery is able to serve its intended goal, it is important to compare its performance to that of an existing test of fine motor control[20]. Therefore, the reproducibility on the tracing and drawing tasks and the modified Fitts' task was compared to the reproducibility of the Purdue pegboard test, since both tests measure aspects of fine motor control. Over the years, the Purdue pegboard test has been used in neuropsychological assessments and rehabilitation contexts[29] and has been shown to be reliable[29–31].

To summarize, the present study investigated the reproducibility of several tracing tasks and a modified Fitts' task. The reproducibility of these graphical tasks was compared to the reproducibility of an independent measure for manual dexterity, the Purdue pegboard task. Additionally, the influence of age on all tasks was investigated.

4.3. Methods

4.3.1. Participants

Thirty-six healthy volunteers, recruited from the general population, participated in this study. The participants were divided over three age-groups of 12 participants each: 20–29 years (mean age 26.3, sd 2.5, 7 males), 30–55 years (mean age 42.0, sd 6.4, 8 males) and 56–75 years (mean age 64.7, sd 6.2, 8 males). All participants signed informed consent and completed the tasks twice with one week in between. Exclusion criteria were a history of epileptic seizures, head injury, neurological or motor disorders, the use of medication affecting movement, or a low (< 26) score on the Mini Mental State Examination (MMSE)). The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen.

4.3.2. Experimental design

Participants were seated in front of a table in a comfortable position to write. A digital tablet (ASUS Eee Slate EP121) and custom made digital pen were used. The position of the pen-tip on the tablet during movement was recorded at a sampling frequency of 200 Hz, using custom developed software. The pen had a wired connection to the tablet. Participants performed eight tasks (see below) with the digitizer pen on the tablet. Additionally, participants performed the Purdue pegboard test. The examiner was seated behind an operator computer to start and stop the recordings. The examiner also determined whether the participants executed the tasks correctly. If a task was executed incorrectly, the recording was stopped and restarted after re-instruction. An example of incorrect task execution would be moving the pen in the wrong direction or starting the task before recording was started.

4.3.3. Tasks

Each participant performed eight tasks in the same order to limit variability in task results. Participants were instructed to start the task at a signal of the examiner and to perform the tasks at a comfortable speed, allowing them to draw as smoothly as individually possible. The participants traced geometric shapes; a circle, a spiral and a zigzag figure (see Figure 4.1). Templates of these shapes were shown on the tablet.

- **Circle tracing:** In this task, participants had to continuously trace a circle ten times in a clockwise direction starting from the 12 o'clock position (see Figure 4.1).
- **Spiral tracing:** In this task the participants traced a spiral (see Figure 4.1) clockwise from inside to outside. Each participant performed ten consecutive spiral tracing trials.
- **ZigZag tracing and drawing:** In this task participants had to trace and draw a zigzag figure (four repeats of a right up line followed by a right down line, see Figure 4.1) from left to right and right to left without lifting the pen. The task consisted of four different conditions and participants performed ten consecutive zigzag drawing trials in each condition. In the first condition (zigzag tracing) participants had to trace the zigzag figure template, which was presented on the tablet. The second condition (zigzag copying) consisted of drawing the zigzag figure with the zigzag figure template provided on paper on the table above the tablet. In the third condition (zigzag blinded) participants had to draw the zigzag figure with their eyes closed and in the fourth condition (zigzag rotated) they had to draw the zigzag figure vertically, while the zigzag figure template was presented horizontally on paper on the table above the tablet.
- **Modified Fitts' task:** The next task was similar to Fitts' original task and adapted for use with the pen and tablet. Two targets, both filled circles, were shown on an imaginary horizontal line in the middle of the tablet. Participants were asked to touch the targets with the pen-tip alternately as fast and as accurately as possible during 20 seconds. In eight subtasks the difficulty of the tasks were altered by varying the distance between targets and varying the diameter of the targets. The varying distances and diameters were chosen according to the dimensions of the tablet. In the first four subtasks (1 to 4), the distance between the centre of the targets was

kept constant at 7 cm, while the diameter of the targets was increased (0.7, 1.3, 1.9, 2.5 cm). In another four subtasks (5 to 8), the distance between the centre of the targets was kept constant at 20 cm, while the diameter of the targets was increased (0.7, 1.3, 1.9, 2.5 cm).

- **Purdue pegboard test:** The Purdue pegboard test employs a board, pins, collars and washers. The board contains two vertically oriented parallel rows with 25 holes in each row and the pins, collars and washers are located in reservoirs at the top of the board. Four subtests were performed. In the first three subtests the participant was instructed to place as many pins as possible in the holes within 30 seconds, first with the dominant hand, then with the other hand and finally with both hands simultaneously. In the last subtest (assembly) the participant used alternate hands to make as many assemblies as possible within 60 seconds. An assembly consisted of a pin, washer, collar and a second washer. In accordance with the instructions, the participants were allowed to practice before each subtest[29].

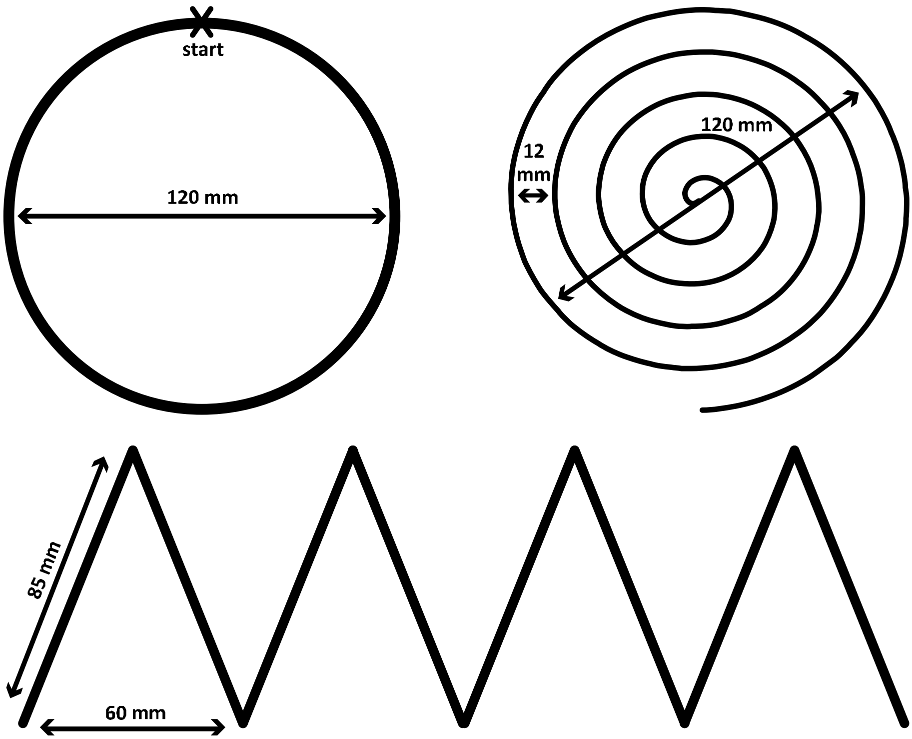

Figure 4.1. Templates and their dimensions for the tracing and drawing tasks: a circle, spiral and zigzag figure.

4.3.4. Data analysis

The drawing and tracing tasks were analysed using custom made scripts in Matlab 7.4.0 (R2007a). Since movement time (MT) was an important measure of speed to distinguish PD patients from HC participants[5], in the present study we also calculated mean MT per repetition for the circle, spiral and zigzag tracing and drawing tasks. Additionally, we calculated the mean deviation (in mm) from the template for the circle, spiral and zigzag tracing tasks (mean error) as a measure of accuracy.

For Fitts' task movement time (MT) and accuracy (Acc) were calculated as well to investigate performance on the task. Movement time was determined as the average time needed to move the pen from one target to the other. To determine accuracy, first each touch was scored according to the distance between the outline of the target and the touch location. The score ranged from 0–1, where a score of 1 was given to the touches inside the target and a score of 0 was given to the touches for which the distance to the outline of the target was more than the radius of the target. The score of the remaining touches was calculated by linear interpolation of the scores between 0 and 1 with distance. Finally, accuracy was expressed as the total accuracy, by summing up the accuracy scores of all touches.

Finally performance on the Purdue pegboard test was analysed. The score on the first two subtests was equal to the number of pins inserted in the holes within 30 seconds. The score on the third subtest equalled the number of pairs of pins inserted in the holes and the assembly score equalled the sum of the number of assembled parts. Also a sum-score was computed by adding the scores obtained in the first three subtests (right hand + left hand + both hands)[29]. All measures were determined for the first and second measurement.

4.3.5. Statistical analysis

<u>Reproducibility</u>

Statistical analyses were conducted using SPSS 20.0.0.1. Since the goal was to investigate the reproducibility of the tasks, the scores on the first and second measurement day were compared for each task and subtask. The degree of agreement between the two measurements for all tasks was determined by the intraclass correlation coefficient (ICC) for the whole group as well as for each age-group separately. In this study we used ICC(C,1) as defined by McGraw and Wong[32] (see Appendix 3 for the estimation of ICC(C,1)). The ICC was calculated in Matlab 7.4.0 (R2007a) and ranges from 0 to 1. According to Andresen[33] an

ICC between 0 and 0.40 signifies poor reliability, between 0.40 and 0.74 moderate reliability and an ICC between 0.75 and 1.00 signifies excellent reliability.

Differences between measurement days and age-groups

Before testing whether differences between measurement days and between age-groups for all measures were significant, a Shapiro-Wilk test was used to test whether the data was normally distributed. If normally distributed, a repeated measures ANOVA with between subject factor 'group' (3 levels; younger, middle-aged and older) and within subject factor 'time' (2 levels; measurement day 1 and day 2) was used to test whether differences between measurement days and age-groups were significant. Otherwise, a Wilcoxon signed rank test was used to test whether differences between the two measurement days were significant. If the data of each of the three age-groups was normally distributed, a one-way ANOVA was used to test whether significant differences between age-groups were present, followed by a post-hoc Bonferroni to determine which groups were significantly different from each other. Otherwise, a Kruskal-Wallis test was used to test whether age-groups were significantly different, with post-hoc Mann Whitney U tests to determine which groups differed significantly from each other. Additionally, difference scores between the two days were calculated for all measures. If the difference scores were normally distributed according to the Shapiro-Wilk test, a one-way ANOVA was used, otherwise a Kruskal-Wallis test was used to investigate whether the age-groups were significantly different from each other. Post-hoc, Mann Whitney U tests were used to determine which groups were significantly different from each other.

4.4. Results

All participants (n=36; mean age: 44.3; sd: 16.8; 23 male; 13 female) completed each of the tracing and drawing tasks and the Purdue pegboard test twice with exactly one week in between. The two measurements were performed at approximately the same time, but at least within a range of three hours on both days. For one participant the MT of the zigzag rotated task on day 1 was missing because of technical problems, so this participant was excluded from further analysis on this task.

Reproducibility

Mean MT and mean error for each of the tracing and drawing tasks are given in Table 4.1 and 4.2 for the total group and for each of the age-groups. Agreement between the first and second measurement for the total group was high for mean MT on the circle, spiral and zigzag tasks (circle tracing: ICC=0.78; spiral tracing: ICC=0.84; zigzag tracing: ICC=0.93; zigzag copying: ICC=0.92; zigzag blinded: ICC=0.91; zigzag rotated: ICC=0.93). In general, mean MT decreased at the second measurement day compared to the first. Agreement for the total group between the first and second measurement for mean error was high for spiral and zigzag tracing (ICC=0.87 and ICC=0.87) and moderate for circle tracing (ICC=0.47).

The descriptive values for MT and Acc for the modified Fitts' task are shown in Tables 4.3 and 4.4, for the total group, as well as for each age-group separately. In general, accuracy increased with an increase in target size and with a decrease in distance between targets. On average, the time to move between targets increased with an increase in distance between targets and decreased with an increase in target size. Additionally, high accuracy was associated with slower movements and vice versa. Agreement between the two measurement days for time was moderate for subtask 1 (ICC=0.72) and high for the other subtasks (all ICC>0.75, see Table 4.3). Agreement between the two measurement days for accuracy was high for subtasks 3 and 7 (ICC=0.82 and ICC=0.80) and moderate for subtask 1, 2, 4−6 and 8 (ICC=0.58, ICC=0.73, ICC=0.74, ICC=0.50, ICC=0.71 and ICC=0.70, respectively, see Table 4.4).

Mean values for the scores on the Purdue pegboard task are shown in Table 4.5. Agreement between the first and second measurement was high for the sum-score (ICC=0.86) and the assembly score (ICC=0.91) (see Table 4.4). The scores for the right hand resulted in an ICC of 0.56, for the left hand in an ICC of 0.80 and for both hands in an ICC of 0.78 (see Table 4.4). All five scores increased slightly on the second measurement day compared to the first. Overall, ICC for most of the measures increased with an increase in age.

<u>Differences between measurement days and age-groups</u>

For the MT on the tracing and drawing tasks a Wilcoxon signed rank test showed a significant difference between measurement days (all p<0.05, see Table 4.1). MT on all tracing tasks was significantly lower on day 2 compared to day 1. There was a significant difference between the age-groups for some of the MT measures on the tracing tasks (indicated in Table 4.1). Post-hoc tests showed that the older group had significantly higher MT, compared to the younger and middle-aged group (p<0.05).

For the mean error on the zigzag tracing task a repeated measures ANOVA showed that mean error was significantly higher on the measurement day 2 compared to day 1 (see Table 4.2). There was also a significant group effect, for which the post-hoc tests showed that the middle-aged group had a significantly higher mean error than the older group ($p<0.05$). No significant interaction effect was found. The mean error on the spiral tracing task was also significantly higher on day 2 compared to day 1 according to a Wilcoxon signed rank test ($p<0.05$, see Table 4.2). No significant difference between the measurement days was found for the mean error on the circle tracing task. There were no significant differences between the groups for the mean error on the circle and spiral tracing task.

For the movement time on subtasks 5, 6 and 7 of the modified Fitt's task a repeated measures ANOVA showed that the average time needed to move between the targets was significantly lower on day 2 compared to day 1 (all $p<0.05$, see Table 4.3). There was also a significant group effect and post-hoc tests showed that the older group needed significantly more time to move between targets compared to the younger and middle-aged group ($p<0.05$). No significant interaction effect was found.

For the movement time on subtask 1,2,3 and 8 of the modified Fitts' task, a Wilcoxon signed rank test showed that movement time was significantly lower on day 2 compared to day 1 (all $p<0.05$, see Table 4.3). For subtask 4 no significant difference was found between the movement time on day 1 and day 2. For subtasks 1,2,3,4 and 8, the age-groups were significantly different (indicated in Table 4.3) and post-hoc tests showed that the older group needed significantly more time to move between the targets than the younger and middle-aged groups ($p<0.05$).

For the total accuracy on subtasks 1,2,3,6,7 and 8 of the modified Fitts' task, a repeated measures ANOVA showed that total accuracy was significantly higher on day 2 compared to day 1 (see Table 4.4). For all subtasks, except subtask 7 there was a significant difference between the groups. Post-hoc tests showed that the older group was significantly less accurate than the younger and middle-aged group for subtask 1, 4 and 5, for subtask 2 the middle-aged group was significantly more accurate than both the younger and older group and for subtask 3 and 8 the older group was significantly less accurate then the middle-aged group (all $p<0.05$). There was no significant interaction effect.

For all the time and accuracy measures for both tracing and the modified Fitts' task the difference scores between day 1 and day 2 did not show significant differences between the age-groups.

For the scores on the Purdue pegboard test a repeated measures ANOVA showed that there was a main effect of time ($p<0.05$, see Table 4.5) for all scores, except for the score with both hands. The scores on day 2 were significantly higher compared to day 1. Also a main effect of group was found for almost all scores, except for the score with the left hand. Post-hoc tests showed that the older group scores significantly lower than the younger and middle-aged group ($p<0.05$, see Table 4.5). There were no significant interaction effects.

Table 4.1. Statistic measures for the movement time (MT) on the tracing tasks. Intraclass correlation coefficients (ICC) and mean movement time per repetition were displayed for the whole group as well as for the three groups separately. Also the results of the Wilcoxon signed rank tests are shown in this table.

	MT(s) Day 1 (Mean(sd))	MT(s) Day 2 (Mean(sd))	ICC	Z-value	p-value
Circle Tracing					
Group I (20–29 years)	3.6 (1.3)	2.3 (0.8)	0.34		
Group II (30–55 years)	3.1 (1.2)	2.3 (1.1)	0.90		
Group III (56–75 years)	4.5 (2.6)	3.1 (1.8)[a]	0.81		
Total	3.2 (2.0)[a]	2.5 (1.5)[a]	0.78	−4.07	0.00
Spiral tracing					
Group I (20–29 years)	7.4 (1.9)	6.0 (1.2)	0.57		
Group II (30–55 years)	7.5 (2.4)	5.9 (1.7)	0.80		
Group III (56–75 years)	8.8 (3.6)[a]	9.1 (4.4)	0.83		
Total	7.9 (3.1)[a]	6.6 (2.5)[a,b]	0.84	−4.15	0.00
ZigZag tracing					
Group I (20–29 years)	7.4 (1.6)	6.3 (1.5)[a]	0.72		
Group II (30–55 years)	7.5 (2.4)	5.9 (1.3)	0.84		
Group III (56–75 years)	10.2 (6.5)[a]	8.6 (6.5)[a]	0.94		
Total	7.6 (3.7)[a]	6.3 (2.8)[a]	0.93	−4.12	0.00
ZigZag copying					
Group I (20–29 years)	6.4 (0.9)	5.5 (1.7)[a]	0.52		
Group II (30–55 years)	6.4 (1.4)	5.5 (1.0)	0.89		
Group III (56–75 years)	8.4 (6.9)[a]	7.0 (5.1)[a]	0.92		
Total	6.6 (2.2)[a,c]	5.6 (1.9)[a,c]	0.92	−4.60	0.00
ZigZag blinded					
Group I (20–29 years)	6.3 (0.8)	5.7 (1.1)[a]	0.37		
Group II (30–55 years)	6.1 (1.2)	5.6 (0.8)	0.84		
Group III (56–75 years)	9.4 (3.8)	8.1 (3.6)	0.94		
Total	6.6 (2.0)[a,b]	5.8 (1.6)[a]	0.91	−4.04	0.00
ZigZag rotated					
Group I (20–29 years)	5.9 (0.8)	5.3 (0.9)	0.61		
Group II (30–55 years)	6.1 (1.1)	5.4 (0.7)	0.76		
Group III (56–75 years)	9.4 (3.2)	7.9 (3.1)	0.95		
Total	6.7 (1.9)[a,b]	5.6 (1.4)[a,b]	0.93	−4.50	0.00

a) Median(iqr) values are displayed; b) Significantly different between the groups according to a one-way ANOVA analysis ($p<0.05$); c) These measures were significantly different between the groups according to a Kruskall Wallis test ($p<0.05$); sd=standard deviation; iqr=interquartile range; Z=Wilcoxon signed rank test for differences between measurement days.

Table 4.2. Statistic measures for the mean error on the tracing tasks. Intraclass correlation coefficients (ICC) and mean error (mean deviation (mm) from template per repetition) are displayed for the whole group as well as for the three groups separately. Also the results of the statistical analysis for differences between measurement days and between groups are shown.

	Error(mm) Day 1 (Mean(sd))	**Error(mm) Day 2 (Mean(sd))**	**ICC**	**Test statistic**	**p-value**
Circle tracing					
Group I (20–29 years)	2.11 (0.39)[a]	2.10 (1.29)[a]	0.27		
Group II (30–55 years)	2.31 (0.61)	2.24 (1.39)[a]	0.56		
Group III (56–75 years)	1.97 (1.83)[a]	2.08 (1.33)[a]	0.65		
Total	2.11 (0.84)[a]	2.12 (1.38)[a]	0.47	−1.05[b]	0.29
Spiral tracing					
Group I (20–29 years)	2.41 (0.17)[a]	2.49 (0.39)[a]	0.75		
Group II (30–55 years)	2.47 (0.54)	2.62 (0.61)	0.87		
Group III (56–75 years)	2.58 (0.73)	2.69 (0.91)	0.92		
Total	2.41 (0.41)[a]	2.49 (0.63)[a]	0.87	−2.29[b]	0.02
ZigZag tracing					
Group I (20–29 years)	2.08 (0.41)	2.36 (0.54)	0.68		
Group II (30–55 years)	2.44 (0.84)	2.78 (0.77)	0.96		
Group III (56–75 years)	1.81 (0.65)	2.01 (0.83)	0.79		
Total	2.11 (0.69)[d]	2.38 (0.77)[d]	0.87	18.69[c]	0.00

a) Median(iqr) values are displayed; b) The Z-values as a result of the Wilcoxon signed rank test for differences between measurement days; c) The F-value as a result of the main effect of time of the repeated measures ANOVA; d. The age-groups were significantly different according to a repeated measures ANOVA ($p<0.05$).

Table 4.3. Statistical results for time on the modified Fitts task. Intraclass correlation coefficients (ICC) and average time needed to move from one target to the other (time(s)) were calculated for the whole group as well as for the three groups separately.

	Time(s) Day 1 (Mean(sd))	Time(s) Day 2 (Mean(sd))	ICC	Test-statistic	P-value
Condition 1					
Group I (20–29 years)	0.39 (0.07)	0.35 (0.07)	0.11		
Group II (30–55 years)	0.36 (0.06)	0.32 (0.06)	0.32		
Group III (56–75 years)	0.51 (0.14)	0.44 (0.13)	0.83		
Total	0.40 (0.13)[a,e]	0.37 (0.10) [e]	0.72	−3.03[b]	0.00
Condition 2					
Group I (20–29 years)	0.32 (0.06)	0.30 (0.05)	0.23		
Group II (30–55 years)	0.29 (0.05)	0.26 (0.04)	0.62		
Group III (56–75 years)	0.42 (0.11)	0.38 (0.10)	0.88		
Total	0.32 (0.11)[a,e]	0.29 (0.09)[a,e]	0.81	−3.10 [b]	0.00
Condition 3					
Group I (20–29 years)	0.27 (0.03)	0.26 (0.04)	0.29		
Group II (30–55 years)	0.24 (0.05)[a]	0.21 (0.04)[a]	0.72		
Group III (56–75 years)	0.35 (0.10)	0.32 (0.09)	0.89		
Total	0.26 (0.08)[a,f]	0.25 (0.07)[a,f]	0.87	−3.1 [b]	0.00
Condition 4					
Group I (20–29 years)	0.24 (0.02)	0.23 (0.03)	0.49		
Group II (30–55 years)	0.22 (0.02)	0.21 (0.03)	0.69		
Group III (56–75 years)	0.30 (0.08)	0.30 (0.07	0.77		
Total	0.24 (0.05)[a,e]	0.23 (0.05)[a,e]	0.83	−1.26 [b]	0.21
Condition 5					
Group I (20–29 years)	0.54 (0.08)	0.54 (0.09)	0.44		
Group II (30–55 years)	0.54 (0.11)	0.49 (0.11)	0.76		
Group III (56–75 years)	0.68 (0.12)	0.61 (0.17)	0.87		
Total	0.59 (0.12)[d]	0.55 (0.13) [d]	0.79	13.36 [c]	0.00
Condition 6					
Group I (20–29 years)	0.49 (0.05)	0.46 (0.07)	0.51		
Group II (30–55 years)	0.47 (0.09)	0.43 (0.10)	0.70		
Group III (56–75 years)	0.58 (0.10)	0.53 (0.15)	0.84		
Total	0.51 (0.09) [d]	0.47 (0.12) [d]	0.80	10.24[c]	0.00
Condition 7					
Group I (20–29 years)	0.43 (0.05)	0.39 (0.06)	0.57		
Group II (30–55 years)	0.40 (0.06)	0.37 (0.08)	0.74		
Group III (56–75 years)	0.51 (0.09)	0.45 (0.13)	0.86		
Total	0.44 (0.08) [d]	0.40 (0.10) [d]	0.83	8.14[c]	0.01
Condition 8					
Group I (20–29 years)	0.37 (0.05)	0.35 (0.05)	0.56		
Group II (30–55 years)	0.36 (0.05)	0.33 (0.06)	0.68		
Group III (56–75 years)	0.46 (0.17)[a]	0.42 (0.10)	0.88		
Total	0.38 (0.09)[a,f]	0.34 (0.12)[a,e]	0.84	−3.41[b]	0.00

a) Median(iqr) values are displayed; b) Z-values as a result of the Wilcoxon signed rank test for differences between measurement days; c) F-values as a result of the main effect of time of the repeated measures ANOVA; d) The age-groups were significantly different according to a repeated measures ANOVA ($p<0.05$); e) The age-groups were significantly different according to a one-way ANOVA ($p<0.05$); f) The age-groups were significantly different according to a Kruskall Wallis test ($p<0.05$). sd=standard deviation; iqr=interquartile range.

Table 4.4. Statistical results for the accuracy measure of the modified Fitts' task. Intraclass correlation coefficients (ICC) and total accuracy (Acc; sum of the accuracy scores of all touches) were calculated for the whole group as well as for the three groups separately.

	Acc Day 1 (Mean(sd))	Acc Day 2 (Mean(sd))	ICC	$F(1,33)^b$	p-value
Condition 1					
Group I (20–29 years)	46.0 (6.0)	50.7 (6.4)	0.00		
Group II (30–55 years)	48.2 (7.1)	54.8 (8.9)	0.62		
Group III (56–75 years)	39.0 (8.5)	39.0 (14.1)[a]	0.69		
Total	44.4 (8.1)[c]	49.5 (9.5)[c]	0.57	13.49	0.00
Condition 2					
Group I (20–29 years)	59.4 (9.5)	60.6 (9.9)[a]	0.10		
Group II (30–55 years)	69.4 (11.7)	75.8 (10.1)	0.51		
Group III (56–75 years)	49.3 (11.4)	55.0 (14.8)	0.86		
Total	59.4 (13.5)[c]	64.8 (14.4)[c]	0.73	9.34	0.00
Condition 3					
Group I (20–29 years)	73.0 (9.2)	77.4 (12.2)	0.20		
Group II (30–55 years)	83.4 (12.3)	90.0 (10.1)	0.77		
Group III (56–75 years)	60.2 (16.8)	65.9 (17.3)	0.90		
Total	72.2 (16.0)[c]	77.7 (16.5)[c]	0.82	10.72	0.00
Condition 4					
Group I (20–29 years)	83.3 (7.8)	87.6 (9.5)	0.33		
Group II (30–55 years)	91.9 (9.6)	95.4 (12.0)	0.55		
Group III (56–75 years)	70.3 (17.3)	70.0 (15.1)	0.67		
Total	81.8 (14.9)[c]	84.3 (16.1)[c]	0.74	1.79	0.19
Condition 5					
Group I (20–29 years)	32.2 (2.6)	32.4 (3.8)	0.39		
Group II (30–55 years)	32.1 (4.4)	33.9 (4.9)	0.60		
Group III (56–75 years)	28.0 (3.8)	28.8 (5.9)	0.21		
Total	30.8 (4.1)[c]	31.7 (5.3)[c]	0.50	1.38	0.25
Condition 6					
Group I (20–29 years)	40.5 (3.9)	41.5 (4.0)	0.55		
Group II (30–55 years)	41.2 (6.7)	42.5 (6.1)	0.48		
Group III (56–75 years)	35.0 (5.9)	37.9 (8.4)	0.86		
Total	38.9 (6.1)[c]	40.7 (6.6)[c]	0.71	4.66	0.04
Condition 7					
Group I (20–29 years)	47.1 (4.7)	49.2 (5.8)[a]	0.63		
Group II (30–55 years)	49.3 (7.4)	50.7 (7.1)	0.69		
Group III (56–75 years)	41.1 (7.8)	45.6 (10.8)	0.90		
Total	45.8 (7.5)	49.0 (8.8)	0.80	13.42	0.00
Condition 8					
Group I (20–29 years)	54.2 (6.9)	58.3 (8.4)	0.58		
Group II (30–55 years)	56.0 (6.5)	60.8 (9.6)	0.38		
Group III (56–75 years)	44.4 (18.5)[a]	48.4 (9.2)	0.79		
Total	52.2 (8.9)[c]	55.8 (10.4)[c]	0.70	8.05	0.00

a) Median (iqr) are displayed; b) F-values as a result of the main effect of time of the repeated measures ANOVA; c) The age-groups were significantly different according to a repeated measures ANOVA ($p<0.05$). sd=standard deviation; iqr=interquartile range.

Table 4.5. Descriptive and statistic measures for the scores of the Purdue Pegboard task. Intraclass correlation coefficients (ICC) and mean and standard deviation (sd) of the scores are displayed for the whole group as well as for the three groups separately. Also the results of the repeated measures ANOVA are shown in this table, which indicate whether significant differences were present between the two measurement days and between the three groups.

	Right hand score[a]	**Left hand score[a]**	**Both hands score[a]**	**Sum-score[b]**	**Assembly score[c]**
	Day 1				
Group I (20–29 years)	14.9 (1.8)	13.2 (1.0)	11.2 (1.0)	39.3 (3.0)	37.7 (4.4)
Group II (30–55 years)	14.4 (2.0)	13.9 (2.1)	11.6 (2.2)	39.9 (5.6)	34.3 (8.4)
Group III (56–75 years)	13.7 (1.3)	12.8 (1.3)	9.8 (1.3)	36.3 (2.6)	29.2 (4.0)
Total	14.3 (1.8)	13.3 (1.6)	10.9 (1.7)	38.5 (4.2)	33.7 (6.8)
	Day 2				
Group I (20–29 years)	15.9 (1.2)	13.9 (1.1)	11.7 (0.9)	41.5 (2.7)	39.4 (4.1)
Group II (30–55 years)	15.1 (1.9)	14.9 (1.9)	11.5 (2.2)	41.5 (5.6)	34.9 (7.2)
Group III (56–75 years)	14.5 (1.4)	13.5 (1.7)	10.3 (1.3)	38.3 (3.5)	29.8 (4.8)
Total	15.2 (1.6)	14.1 (1.7)	11.1 (1.6)	40.4 (4.3)	34.7 (6.7)
	ICC				
Group I (20–29 years)	0.27	0.58	0.43	0.59	0.73
Group II (30–55 years)	0.81	0.93	0.88	0.93	0.94
Group III (56–75 years)	0.19	0.59	0.59	0.76	0.80
Total	0.56	0.80	0.78	0.86	0.91
	Repeated Measures ANOVA				
Main effect of Time	F(1,32)=9.027; p=0.005	F(1,32)=18.758; p=0.000	F(1,32)=1.978; p=0.169	F(1,32)=22.995; p=0.000	F(1,32)=8.349; p=0.007
Main effect of Group	F(2,32)=9.167; p=0.001	F(2,32)=1.137; p=0.333	F(2,32)=8.013; p=0.002	F(2,32)=6.753; p=0.004	F(2,32)=7.883); p=0.002
Group 1 vs 2	p=1.000		p=1.000	p=1.000	p=1.000
Group 1 vs 3	p=0.001		p=0.003	p=0.005	p=0.003
Group 2 vs 3	p=0.009		p=0.007	p=0.018	p=0.009

a) Number of pins inserted in the holes with the right hand; left hand and both hands simultaneously; b) Sum-score=Right hand score + Left hand score + Both hands score; c) Assembly score=Number of assembled pins, collars and washers.

4.5. Discussion

The aim of this study was to investigate the reproducibility of a set of standardized graphical tasks using a digital pen and tablet. Overall, the performance measures derived for the tasks showed moderate to excellent test-retest reliability. Additionally, this study showed that in general test-retest reliability increased with age.

We showed that reproducibility on this set of graphical tasks, which measures aspects of fine motor control, was similar to the reproducibility of an already validated fine motor control task, the Purdue pegboard test. This suggests that the set of graphical tasks studied here provides a reliable method to measure aspects of fine motor control. Mean MT and mean error per repetition for the tracing tasks were well reproducible, in line with previous studies which also reported high test-retest reliability for MT on a circle drawing task and some handwriting tasks[17] and mean drawing error on a tracing task[21], although Pearson or Spearman correlation coefficients were used in those studies to assess reproducibility[17]. In our study, participants were significantly faster on the second measurement day compared to the first, which suggests a learning effect. According to Longstaff and Heath[9] handwriting and drawing are overlearnt skills and are not expected to considerably improve or deteriorate over time. However, a possible learning effect could be stronger in simple tasks compared to more complex tasks. It is, for example, easier to increase speed on a simple circle tracing task than on a spiral tracing task, as the spiral tracing task requires more accuracy. In the present study, the most complex tracing and drawing task – the zigzag task – indeed showed better reproducibility than the simpler circle and spiral tracing tasks. This confirms that there is a smaller learning effect for the zigzag task. In addition, the mean error for the spiral and zigzag tracing tasks showed higher reproducibility than the mean error for the circle tracing task. Similar results were found for the Purdue pegboard test, for which reproducibility was also better on the complex task than on the simple tasks. This finding suggests that complex tasks are more reliable than simple tasks to assess fine motor control, since learning effects between two measurements are smaller.

We showed that reproducibility for movement time on a modified Fitts' task was good for all subtasks, and reproducibility was moderate to good for the accuracy measure for all subtasks. Movement time seems more reliable than accuracy as a measure of performance on a modified Fitts' task. However, accuracy was least reproducible for the most complex subtasks of the modified

Fitts' task, which indicates that these subtasks might be too difficult and could be removed from the task-battery.

Movement time on all tasks increased with age, consistent with previous studies[17,34]. However, the results of the older group were generally more reproducible than the results of the other two groups, which suggests that performance (in terms of speed) in the older group is more stable over time. An explanation may be that older adults payed more attention to performing the tasks correctly while the younger and middle-aged groups were more focused on finishing the tasks quickly, which may cause a larger learning effect in the latter groups. High reproducibility of the speed at which fine motor control tasks are executed in the older group indicates that these tasks might be particularly suited for application in movement disorders such as PD, which is typically diagnosed in people older than 60 years, but less suited in movement disorders which are diagnosed at very young or across all ages. On the contrary, mean error for the spiral and zigzag tracing tasks were highly reproducible and did not significantly differ between age-groups, suggesting that drawing error might be more suited for diagnostic or monitoring applications in movement disorders that occur across all ages.

Graphical tasks have previously been proposed as an aid in the diagnostic work-up of movement disorders, since differences in performance on these tasks have been found between patients with movement disorders and healthy controls[4–7,10,18,19] and between patients with different movement disorders[8]. However, none of these studies included reproducibility testing, which is a necessary step before introducing a test in clinical practice[20]. In the present study we showed high reproducibility of a set of graphical tasks executed with the same pen and tablet, and for the tracing tasks we have already shown differences between PD patients and HC participants[5]. Further testing of the set of graphical tasks used in the present study is still needed with additional analysis, to show that PD patients can be distinguished from patients with other movement disorders, such as essential tremor. To investigate whether these graphical tasks are suitable to evaluate treatment effects, the measurements should be repeated with additional analysis in patients before and after being treated with medication or other therapies. Additionally, performance on such graphical tasks should be validated against current gold standards in movement disorders, such as the Unified Parkinson's Disease Rating Scale (UPDRS).

4.6. Conclusions

To conclude, this study shows that a set of graphical tasks, which measure fine motor control, has high reproducibility and that reproducibility is similar to the reproducibility of another fine motor control task, the Purdue pegboard task. We propose that more complex tasks, such as zigzag drawing and Fitts' task are more suitable for clinical testing, because such tasks are more reliable than simple tasks. The time and accuracy measure of the Fitts' task seem to be specifically useful for testing in movement disorders which are common in older age-groups, such as PD, since these measures showed superior reliability in the older group compared to the younger and middle-aged group.

4.7. References

[1] Blank, R., Miller, V. & Voû, H. Von. Human motor development and hand laterality: a kinematic analysis of drawing movements. *Neurosci. Lett.* **295,** 89–92 (2000).

[2] Zietsma, R. C. Designing a comprehensive system for analysis of handwriting biomechanics in relation to neuromotor control of handwriting (PhD thesis). (University of Strathclyde, Glasgow, 2010). at <http://www.strath.ac.uk/library/>

[3] Jankovic, J. Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry* **79,** 368–376 (2008).

[4] Broderick, M. P., Van Gemmert, A. W. A., Shill, H. a & Stelmach, G. E. Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease. *Exp. Brain Res.* **197,** 223–33 (2009).

[5] Smits, E. J., Tolonen, A. J., Cluitmans, L., Van Gils, M., Conway, B. a., *et al.* Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease. *PLoS One* **9,** 1–8 (2014).

[6] Ponsen, M. M., Daffertshofer, A., Wolters, E. C., Beek, P. J. & Berendse, H. W. Impairment of complex upper limb motor function in de novo Parkinson's disease. *Parkinsonism Relat. Disord.* **14,** 199–204 (2008).

[7] Rosenblum, S., Samuel, M., Zlotnik, S., Erikh, I. & Schlesinger, I. Handwriting as an objective tool for Parkinson's disease diagnosis. *J. Neurol.* **260,** 2357–2361 (2013).

[8] Bajaj, N. P. S., Wang, L., Gontu, V., Grosset, D. G. & Bain, P. G. Accuracy of subjective and objective handwriting assessment for differentiating Parkinson's disease from tremulous subjects without evidence of dopaminergic deficits (SWEDDs): an FP-CIT-validated study. *J. Neurol.* **259,** 2335–2340 (2012).

[9] Longstaff, M. G. & Heath, R. A. Spiral drawing performance as an indicator of fine motor function in people with multiple sclerosis. *Hum. Mov. Sci.* **25,** 474–491 (2006).

[10] Dounskaia, N., Van Gemmert, A. W. A., Leis, B. C. & Stelmach, G. E. Biased wrist and finger coordination in Parkinsonian patients during performance of graphical tasks. *Neuropsychologia* **47,** 2504–2514 (2009).

[11] Keresztényi, Z., Cesari, P., Fazekas, G. & Laczkó, J. The relation of hand and arm configuration variances while tracking geometric figures in Parkinson's disease: aspects for rehabilitation. *Int. J. Rehabil. Res.* **32,** 53–63 (2009).

[12] Bryant, M. S., Rintala, D. H., Lai, E. C. & Protas, E. J. An investigation of two interventions for micrographia in individuals with Parkinson's disease. *Clin. Rehabil.* **24,** 1021–1026 (2010).

[13] Sülzenbrück, S., Hegele, M., Heuer, H. & Rinkenauer, G. Generalized slowing is not that general in older adults: Evidence from a tracing task. *Occup. Ergnomics* **9,** 111–117 (2010).

[14] Sülzenbrück, S., Hegele, M., Rinkenauer, G., Heuer, H., Hegele, M., *et al.* The Death of Handwriting: Secondary Effects of Frequent Computer Use on Basic Motor Skills. *J. Mot.* **43,** 247–251 (2011).

[15] van Drempt, N., McCluskey, A. & Lannin, N. A. Handwriting in healthy people aged 65 years and over. *Aust. Occup. Ther. J.* **58,** 276–286 (2011).

[16] Bashir, M. & Kempf, F. Advanced Biometric Pen System for Recording and Analyzing Handwriting. *J. Signal Process. Syst.* **68,** 75–81 (2011).

[17] Mergl, R., Tigges, P., Schröter, A., Möller, H. J. & Hegerl, U. Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives. *J. Neurosci. Methods* **90,** 157–69 (1999).

[18] Ünlü, A., Brause, R. & Krakow, K. Handwriting Analysis for Diagnosis and Prognosis of Parkinson ' s Disease. *ISBMDA Lect. Notes Comput. Sci.* **4345,** 441–450 (2006).

[19] Saunders-Pullman, R., Derby, C., Stanley, K., Floyd, A., Bressman, S., *et al.* Validity of spiral analysis in early Parkinson's disease. *Mov. Disord.* **23,** 531–537 (2008).

[20] Van den Bruel, A., Cleemput, I., Aertgeerts, B., Ramaekers, D. & Buntinx, F. The evaluation of diagnostic tests: evidence on technical and diagnostic accuracy, impact on patient outcome and cost-effectiveness is needed. *J. Clin. Epidemiol.* **60,** 1116–1122 (2007).

[21] Erasmus, L., Sarno, S., Albrecht, H. & Schwecht, M. Measurement of ataxic symptoms with a graphic tablet : standard values in controls and validity in Multiple Sclerosis patients. *J. Neurosci. Methods* **108,** 25–37 (2001).

[22] Feys, P., Helsen, W., Prinsmel, A., Ilsbroukx, S., Wang, S., *et al.* Digitised spirography as an evaluation tool for intention tremor in multiple sclerosis. *J. Neurosci. Methods* **160,** 309–316 (2007).

[23] Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. 1954. *J. Exp. Psychol. Gen.* **121,** 262–269 (1992).

[24] Jagacinski, R. J. & Monk, D. L. Fitts' Law in two dimensions with hand and head movements. *Journal of motor behavior* **17,** 77–95 (1985).

[25] Langolf, G. D., Chaffin, D. B. & Foulke, J. A. An Investigation of Fitts' Law Using a Wide Range of Movement Amplitudes. *J. Mot. Behav.* **8,** 113–128 (1976).

[26] Christe, B., Burkhard, P. R., Pegna, A. J., Mayer, E. & Hauert, C. Clinical assessment of motor function: a processes oriented instrument based on a speed-accuracy trade-off paradigm. *Behav. Neurol.* **18,** 19–29 (2007).

[27] Sanes, J. N. Information processing deficits in Parkinson's disease during movement. *Neuropsychologia* **23,** 381–392 (1985).

[28] Rand, M. K., Stelmach, G. E. & Bloedel, J. R. Movement accuracy constraints in Parkinson's disease patients. *Neuropsychologia* **38,** 203–212 (2000).

[29] Desrosiers, J., Hébert, R., Bravo, G. & Dutil, E. The Purdue Pegboard Test: normative data for people aged 60 and over. *Disabil. Rehabil.* **17,** 217–224 (1995).

[30] Tiffin, J. & Asher, E. J. The Purdue pegboard; norms and studies of reliability and validity. *J. Appl. Psychol.* **32,** 234–247 (1948).

[31] Müller, T., Schäfer, S., Kuhn, W. & Przuntek, H. Correlation between tapping and inserting of pegs in Parkinson's disease. *Can. J. Neurol. Sci.* **27,** 311–315 (2000).

[32] McGraw, K. O. & Wong, S. P. 'Forming inferences about some intraclass correlations coefficients': Correction. *Psychol. Methods* **1,** 390–390 (1996).

[33] Andresen, E. M. Criteria for assessing the tools of disability outcomes research. *Arch. Phys. Med. Rehabil.* **81,** 15–20 (2000).

[34] Holper, L., Kiper, D. C., Eng, K. & Vuillermot, S. An extended drawing test for the assessment of arm and hand function with a performance invariant for healthy subjects. *J. Neurosci. Methods* **177,** 452–460 (2009).