Musculoskeletal Physiotherapy Sciences

Master Research Project 2018-2019

The reliability of the Drivkraft Wheelchair Manoeuvre Test

Vrije Universiteit Amsterdam
Department of Human Movement Sciences
Qualification: MSc in Musculoskeletal Physiotherapy Sciences
Master Research Project

Author: *Michael Bernardus Gustaaf Zonneveld (2625326)*

Supervisors: *Prof. dr. Thomas Janssen*
Kerstin Wahman, PhD. RPT
Åke Norsten

August 2019

The reliability of the Drivkraft Wheelchair Manoeuvre Test

Michael B.G. Zonneveld

Abstract

Background and purpose: Manual wheelchair users have a potential risk for musculoskeletal overuse injuries. To prevent these overuse injuries wheelchair users need to be taught how to propel a wheelchair efficiently. To test how a person is manoeuvring the wheelchair the Drivkraft Wheelchair Manoeuvre Test (DWMT) was developed. The aim of this study is to determine the reliability components of the DWMT.

Method: The inter- intra-rater and test-retest reliability were determined for the four tasks of the DWMT. Three raters scored the test for the inter- and intra-rater reliability and one rater scored both the test and the retest in a different group of participants. Intraclass correlation coefficients (ICC) and kappa values were calculated.

Results: 30 participants were included for the inter- and intra-rater reliability and 18 were included for the test-retest reliability. The inter-rater reliability showed ICC values that were good to excellent and the kappa was found to be strong to almost perfect. The intra-rater reliability values are excellent for all raters. Test-retest reliability showed perfect kappa values and excellent ICC values.

Conclusion: The DWMT has good to excellent inter- intra-rater and test-retest reliability when sufficient sample sizes are used. Further research can be conducted using the DWMT.

Keywords: Motor skills; Psychometrics; Rehabilitation; Reliability; Reproducibility of results; Wheelchairs;

Introduction

The wheelchair is one of the most common and important rehabilitation devices (1) used by individuals who have limited ability to walk (2). According to the most recent data, 1% of the Swedish population uses a wheelchair, equating to about 150.000 wheelchair users. Of these people, 67% uses a manual wheelchair (3).

Manual wheelchairs are propelled with a pushrim, which has been found to be inefficient (4) and stressful to the musculoskeletal structures of the upper extremity (5, 6). The inefficiency can result in overuse injuries of the shoulder (7-11) and wrist (12-15). This is not only due to using the wheelchair all day, but also due to a range of barriers encountered by manual wheelchair users (16) and having the shoulder as the main weight-bearing joint (17). Trying to overcome these barriers, without having received training in how to do so, increases the potential for acute or long-term injuries (18-20). This in turn has been found to have a negative effect on the perceived quality of life of the manual wheelchair users (21). It is therefore necessary to instruct the manual wheelchair user and focus on teaching them appropriate skills (22-24). This training is usually done in a rehabilitation center by asking the participants to perform tasks in the wheelchair in situations reflecting their needs (25, 26).

One of the methods to train the manual wheelchair user is provided at a wheelchair school in a Swedish rehabilitation center and is called Drivkraft (27). In The Wheelchair School, not only wheelchair skills are trained, but the wheelchair is also adapted to a participant's physical measurements and capabilities, and participants also learn relevant theory about the wheelchair, such as the adjustments that can be made, cleaning, and maintenance. Over a period of five weeks participants will follow the training consisting of three lessons a week for a duration of three hours. The Drivkraft training is therefore a more intense training program than what is regularly encountered in rehabilitation programs (28). The training is given to heterogeneous groups with a maximum of five participants and with attention to individual goals.

The Wheelchair School is based on three main components with subcomponents. These components are: 1) wheelchair adjustments (posture, working position, and mechanics); 2) wheelchair training (basic skills, obstacles, movement patterns, and control of center of gravity); 3) patient education (wheeled mobility). Characteristic of this training is that it is taught by a physical education teacher who is also a peer mentor, who has been continuously adapting and improving the method since 1996.

To be able to study if the wheelchair training component of The Wheelchair School has an effect it is important to be able to measure the wheelchair skills, by using a wheelchair skills test (29). An earlier review found that only three tests included the measurement of rear-wheel balance skills (29), while this is considered to be one of the most important skills to be taught in The Wheelchair School. The more recently developed and studied wheelchair skills tests, such as: the Wheelchair Skills Test 4.2 (30), Wheelchair Circuit Test (31), and the Queensland Evaluation of Wheelchair Skills (QEWS) (32), did include rear-wheel balancing assessments to evaluate wheelchair skills.

These tests, although valid and reliable, are not suitable for The Wheelchair School for a few reasons. The Wheelchair Skills Test 4.2 (33) contains an extensive list of skills that have to be performed and takes approximately 30 minutes to complete. A strength of this test is that it keeps being updated and researched in different patient groups. The skills in this test include a range from the basics such as rolling the wheelchair forward to more advanced skills such as descending stairs. This test includes skills that are not being trained at The Wheelchair School and with a duration of 30 minutes takes too long to administer. The Wheelchair Circuit Test (31) has shown reliability and validity for participants with paraplegia and tetraplegia, however the test showed a ceiling effect for people with paraplegia. Because the Wheelchair Circuit Test was only tested on those with para- and tetraplegia it does not match with the intended heterogeneity of The Wheelchair School. The test also takes more than 30 minutes and was therefore not the most suitable for using in The Wheelchair School.

One of the more suitable tests for the purpose to evaluate The Wheelchair School is the QEWS. The developers of the QEWS were looking for a test that was easily applicable in a hospital setting (32), and encountered similar problems with the existing tests. The researchers had a Delphi panel compiling a priority list of wheelchair skills, which resulted in the items of the QEWS. Although the items used on the QEWS are similar to those that are being trained in The Wheelchair School, the scoring system is too broad by including assistance in performing the skills. The test also includes ascending and descending a kerb of 15.0 cm, which is 5.0 cm higher than the European Standard (34). The test does not differentiate between lower kerbs, which could also be encountered. The final item is a six minute push test, which adds up to the administration time and is considered to measure fitness (35). Last, ceiling effects might be present in the indoor circuit with a focus only on independent completion. Even though the QEWS had its limitations in the administration when compared to the needs for The Wheelchair School, it was used as a base upon which the Drivkraft Wheelchair Manoeuvre Test (DWMT) has been developed. The DWMT is a skills test with a focus on precision, manoeuvrability and rear wheel balancing, with an expected ceiling effect only for the manual wheelchair users with a high skill level. The scoring system used in the DWMT is more raw and representative of what has actually been achieved by a participant instead of an overall score. The rationale for this way of reporting is that it keeps the results easy to interpret and shows what a participant is capable of. Furthermore the DWMT has been developed in such a way that it is quick to perform and can be performed at multiple locations.

Before the DWMT can be used to evaluate the training of The Wheelchair School the reliability needs to be determined. Because the DWMT consists of multiple components that are scored independently of each other the reliability will be determined for each component. The aim of this study is to determine the reliability of the DWMT components by answering the question: what are the inter- intra-rater and test-retest reliability of the components of the DWMT?

Method

Reliability study

Multiple aspects of the reliability of the DWMT were determined, of which a schematic overview can be seen in Figure 1. First the inter-rater reliability together with the intra-rater reliability were determined. This was done by three raters who were all experienced with wheelchair skills and familiarized themselves with the DWMT before the study. To enable the same test to be scored on multiple occasions by multiple raters the DWMT was video recorded. The video recordings were removed of any unnecessary material, such as changing the set-up of the kerbs and the transitioning between different tasks. These edited videos were used to be scored by the raters. The raters were not allowed to replay the video, because that would have given them an advantage over the live rating done by Rater 1. The DWMT was scored in the same way during the live rating and the video ratings. The scoring criteria were thoroughly described in a protocol which can be found in appendix 1. It was aimed to have two weeks between the first and the second rating for the intra-rater reliability.

The second part was to determine the test-retest reliability for which it was of importance to select participants that would not improve anymore by having done the test multiple times (36). It was also important to let there be enough time between the two measurements to not have an adverse effect due to fatigue (37). Since the test was not expected to be exhausting for experienced wheelchair users a minimum of 2 hours between the test and the retest was considered sufficient, but always on the same day. The test-retest reliability participants were scored by the same rater on both occasions.

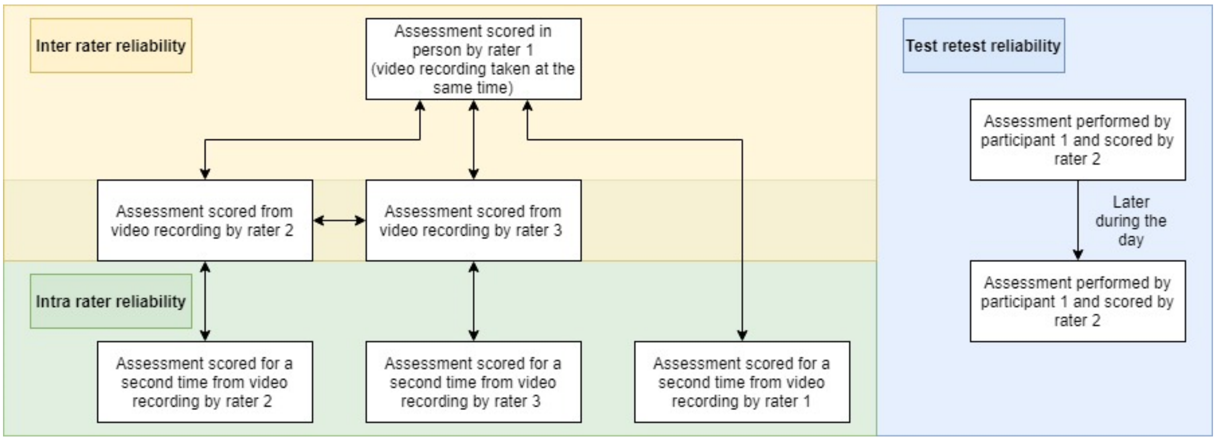


Figure 1: Schematic overview of measuring the inter-rater reliability and intra-rater reliability

Participants

Walter et al. (38) developed formula [1] to calculate the sample sizes for reliability studies. They did some of the more general calculations and published the results in tables. To derive the sample size (k) from these tables the following values were needed: the type I (α) and type II (β) errors, as well as the number of raters or ratings (n) and the intraclass correlation coefficient (ICC) of the null hypothesis (ρ_0) and the alternative hypothesis (ρ_1).

$$k = 1 + \frac{2(U_\alpha + U_\beta)^2\, n}{(\ln C_0)^2 (n-1)} \qquad [1]$$

Bujang & Baharum advised to add an additional 20% to any sample size k to prepare for drop-outs of the study (39). This 20% was taken into account for the final sample size.

Sample size for the inter- and intra-rater reliability:
α = 0.05, β = 0.20, ρ_0 = 0.70, ρ_1 = 0.90 and n = 3, gives k = 12.8, add 20%: k = 15.

Sample size for the test-retest reliability:
α = 0.05, β = 0.20, ρ_0 = 0.70, ρ_1 = 0.90 and n = 2, gives k = 18.4, add 20%: k = 22.

Recruitment of participants was done continuously over a period of three months as the referrals for The Wheelchair School came in. Other participants were inpatients or staff of the rehabilitation centre. To be eligible for participating in the study the participant had to have a neurological disorder and use a manual wheelchair. They also had to speak Swedish or English as those were the languages the instructions of the DWMT could be given in. For the test-retest reliability an additional inclusion criteria applied: the participants had to be stable in their level of wheelchair skills.

The DWMT

The DWMT was administered at the intake or at the first lesson of The Wheelchair School. Participants performed the four tasks to their maximum capability. This was done in their own wheelchair and in case they did not have one a fitting wheelchair with factory settings was provided. The participants were shown by the rater on how to perform the task before attempting it themselves. Task 1 was the only task where the participants were given a trial run before the assessment. All tasks were scored separately and stand on their own. The following tasks were included in the DWMT;

- *Task 1: Fundamental manoeuvring - Precision*
  This task is a slalom course, with both forward and backward manoeuvring. The participant was instructed to complete the course as fast as possible, without making mistakes, and stop at a specified location. The space between two cylinders is set at 89 cm which is the average width of a door in Sweden (40). The time it takes to complete the course is measured and the number of mistakes (e.g. wrong direction or hitting a cylinder) are counted.
- *Task 2: Fundamental manoeuvring – Slope of 1/10.3*
  In this task a slope with a gradient of 1/10.3 has to be ascended and descended. When going down, a slalom is added to the slope and the participant is considered to be precise when no cylinders are hit and he/she stops at the specified location. The two parts are scored as a pass or no pass.
- *Task 3: Control of centre of gravity – Rear-wheel balancing, push and sliding/breaking*
  This task consists of six components, all related to rear-wheel balancing and going up in difficulty. These components are: 1) rear wheel balance for 20 seconds; 2) turn 360 degrees in both directions; 3) high and low balance position over 10 cm; 4) going down a slope in a straight line; 5) going down a slope with a slalom; 6) sliding for 2 meters on a flat surface. Anti-tippers were allowed on the wheelchair, but were checked for their height to make sure the components were possible. All components are scored as a pass or no pass which results in a total score between 0 and 6. However not all the components will be performed by a participant if he/she does not complete a previous component. If component 1 is scored a no pass, the rest of the task will not be assessed. If component 4 is scored a no pass, component 5 and 6 will not be assessed. A more detailed description can be found in the protocol in appendix 1.
- *Task 4: Obstacle technique – Going up and down a kerb*
  This task tests how high the participant can safely ascend and descend a kerb starting at 1.0 cm up to a maximum of 15.0 cm in predefined steps (2.5; 5.0; 7.5; 10.0; 12.5 cm respectively). For this reliability study every ascent and descent correctly performed adds to the total score, ranging from 0 to 14.

While the test was administered by the instructor of The Wheelchair School, a second person was present making the video recording. In case the participant wanted to try a more difficult skill but did not feel safe a third person would ensure safety. The video recording was done according to a standardized protocol, which can be found in appendix 2.

Data analysis

The collected data was entered in SPSS v.23 and the reliability was determined by calculating the ICC together with the standard error of measurement (SEM). In Task 2 the kappa was calculated, since this was the only dichotomous data. The ICC was calculated by using the two-way mixed effects model in SPSS with absolute agreement, because the reliability was calculated for the values the different raters have given (41). To calculate the SEM the variance components were computed by SPSS and these values were used in the formula of the SEM.

For the inter-rater reliability the ICC and SEM or kappa were calculated for the comparison between Rater 1 and 2, Rater 1 and 3, and Rater 2 and 3. To give insight in which of the raters agree with each other and on which level. For the intra-rater reliability every rater was given a value for the ICC and SEM or a kappa based on the scores they have given. The test-retest reliability resulted in a value for the ICC, kappa and SEM for the two tests performed by the same participant. Interpretation of the ICC was according to the guidelines by Koo & Li (41). The kappa was interpreted according to McHugh (42). See Table 1 for the details on the interpretation.

Table 1 interpretation of the reliability values

ICC according to Koo & Li (41)			**Kappa according to McHugh (42)**		
< 0.5	=	poor reliability	< 0.20	=	None
0.5 – 0.74	=	moderate reliability	0.21 – 0.39	=	Minimal
0.75 – 0.9	=	good reliability	0.40 – 0.59	=	Weak
> 0.9	=	excellent reliability	0.60 – 0.79	=	Moderate
			0.80 – 0.90	=	Strong
			> 0.90	=	Almost perfect

ICC = intraclass correlation coefficient

Results

A total of 30 participants were included in the inter- and intra-rater reliability tests and 18 participants were included for the test-retest reliability test. The participants for the test-retest reliability had on average 130 months more experience in using the wheelchair than the participants for the inter- and intra-rater reliability. The participants for the inter- and intra-rater reliability test were equally split in males and females and the test-retest reliability had more males than females. Five participants had just received their first wheelchair, these were all included in the inter- intra-rater reliability tests. An overview of the demographic data of the participants can be seen in Table 2. The test duration was found to be less than 15 minutes on average.

Table 2 demographics

	Inter-intra (N=30)	Test-retest (N=18)
Age (SD), years	53,6 (13,3)	49 (14,4)
Sex		
Male	15	13
Female	15	5
Diagnosis		
SCI	11	11
MS	7	1
CP	1	2
PPS	2	0
Other	9	4
Time in a wheelchair (IQR), months	32 (115)	162 (336)
1st wheelchair	5	0

SD = standard deviation; SCI = spinal cord injury; MS = multiple sclerosis; CP = cerebral palsy; PPS = post-polio syndrome; IQR = inter quartile range

Reliability

The ICC and kappa values for the inter-rater, intra-rater and test-retest reliability can be seen in Table 3, 4 and 5 respectively. The inter-rater reliability ICC values were mostly excellent or in the good-excellent range, with the lowest values of the 95% confidence interval (CI) still reaching into the excellent range (0,779 – 0,945). The SEM showed mostly a high score accuracy, with a difference never greater than ± 1. The SEM sometimes showed a lower accuracy on the mistakes made at Task 1 and the kerb height at Task 4. This was seen when comparing between the raters, but not across all comparisons. The accuracy is visualized in Figure 2 A, B, E and F. The kappa values were found to be strong to almost perfect in Task 2. Figure 2 D shows that there was one participant scored different between Rater 1 and Rater 2. This one participant difference resulted in a kappa of 0,839 (see Table 3) which is interpreted as strong. Figure 2 C shows no deviations between the raters which resulted in a perfect kappa (1,000). The figures for the comparisons between Rater 1 and 3 and between Rater 2 and 3 can be found in Appendix 3. Going down with the slalom in Task 2 showed the largest discrepancy with three participants scored differently between Rater 2 and 3 and between Rater 1 and 3. This resulted in a moderate kappa value. The kappa between Rater 1 and 2 was higher and can be interpreted as strong.

The intra-rater reliability ICC values were found to be excellent for all raters (Table 4), with the lowest values of the 95% CI (0,930 – 0,983) completely in the excellent range. The SEM was similar to the values of the inter-rater reliability, except for Rater 3 who had a SEM of 1,68 for the measured time in Task 1. The kappa values were strong or almost perfect for Task 2. Figure 3 shows the accuracy of Rater 1, which was quite good, although a larger differentiation from the line can be seen in Figure 3 B & F. This deviation is not too big to have an impact on the SEM. The figures of the other raters can be found in appendix 4. The average time between the two ratings was 15 days.

On average there were three hours between the test and the retest. The results, shown in Table 5 and Figure 4, showed excellent values for the ICC and only perfect kappa values (Figure 4 C & D). Except for the mistakes made on Task 1 with an ICC (95%CI) of 0,702 (0,361 – 0,877), which is the lowest ICC found. The SEM of 1,08 is higher than at the inter-rater reliability, but still indicates that the scores were accurate, the visualization of this accuracy is shown in Figure 4 A. Task 3 even had an ICC value of 1,000 where all scores were completely the same in the test and the retest (Figure 4 E).

Discussion

This was the first study done for determining part of the psychometric properties of the DWMT after it was developed. In this study, the inter- intra- and test-retest reliability of the components of the DWMT were examined. The ICC and kappa values found show good to excellent reliability for almost all components. These values were expected to be found, since the QEWS was used as a base and this test has shown similar reliability values (32).

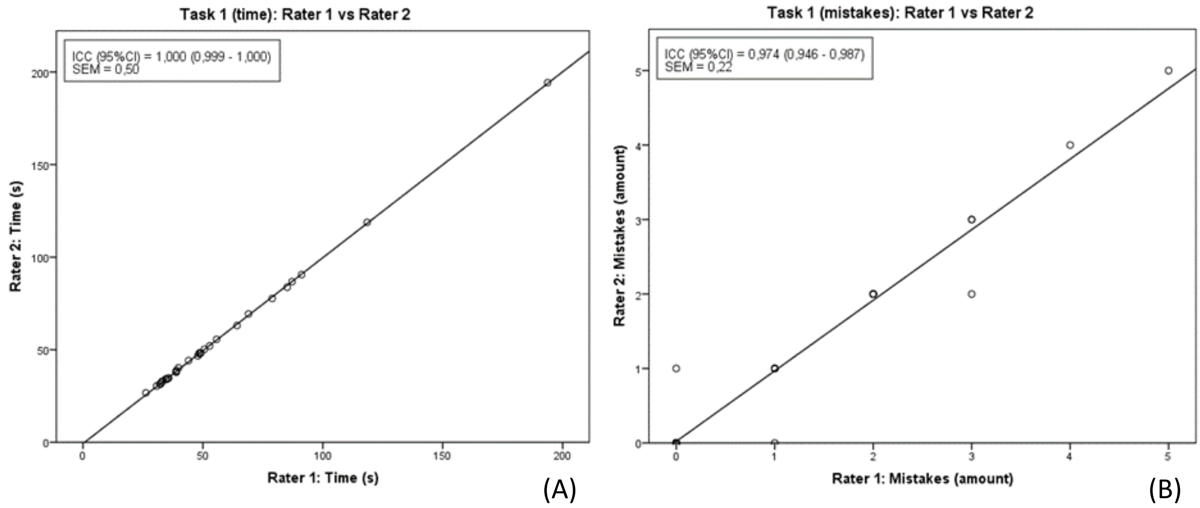
Task 1 was a slalom that was adapted from the QEWS item 1 (32). The course was changed and the time in seconds was noted instead of scoring a time range, additionally the amount of mistakes were noted. The seconds measured show excellent inter- and intra-rater reliability for all raters. However the mistakes counted seemed to have some difficulties. The mistakes show a moderate accuracy for the inter-reliability when comparing Rater 1 and 2 with Rater 3. A reason for this might be an unclarity in the description on how to score in the protocol, since Rater 3 only used the protocol for scoring the participants and was not actively taking part in the development of the scoring criteria. The mistakes were also scored with lower accuracy for the intra-rater reliability. The test-retest reliability had the lowest ICC value on the mistakes of Task 1; ICC(95%CI) of 0,702 (0,361 – 0,877). A possible reason for this is the competitive nature of the participants. Since they knew how it went on the first test they wanted to go faster on the retest in the afternoon. Although not knowing their time from the morning, they went faster but with more mistakes. Therefore counting the mistakes on Task 1 only showed moderate reliability. An option that was considered was to put a time penalty for every mistake made, to have only time in seconds as an outcome of the slalom. This was not done because no single value of time penalty would suit the whole range of scores. Therefore the descriptions of how to decide on a mistake will be checked again in the further development of the DWMT and revisions in the description or definitions might be necessary.

The inter-rater reliability on Task 2 shows quite the range in kappa values, ranging from moderate to almost perfect. By having only the option of scoring a pass or no on the question if the participant was able to complete the task, the scoring relied on the description in the protocol. The requirements for correct interpretation of the kappa have been met (42), and a single difference between two raters has quite the

Table 3 Inter-rater reliability

Inter-rater reliability		Rater 1 vs Rater 2			Rater 1 vs Rater 3			Rater 2 vs Rater 3		
	N	ICC (95% CI)	SEM	κ	ICC (95% CI)	SEM	Κ	ICC (95% CI)	SEM	κ
Task 1										
Time	30	1,000 (0,999 – 1,000)	0,50		1,000 (1,000-1,000)	0,44		1,000 (1,000-1,000)	0,43	
Mistakes	30	0,974 (0,946 – 0,987)	0,22		0,901(0,805-0,952)	0,41		0,888(0,779-0,945)	0,43	
Task 2										
Reaches the top	30			1,000			0,911[a]			0,911[a]
Down with slalom	30			0,839[a]			0,609[c]			0,714[b]
Task 3										
Total score	30	0,994(0,987-0,997)	0,13		0,973(0,938 – 0,988)	0,29		0,969(0,934-0,985)	0,32	
Task 4										
Total score	30	0,983 (0,965 – 0,992)	0,37		0,969(0,937-0,985)	0,50		0,985(0,970-0,993)	0,34	

[a] 1 participant was scored different; [b] 2 participants were scored different; [c] 3 participants were scored different; ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement; κ = kappa



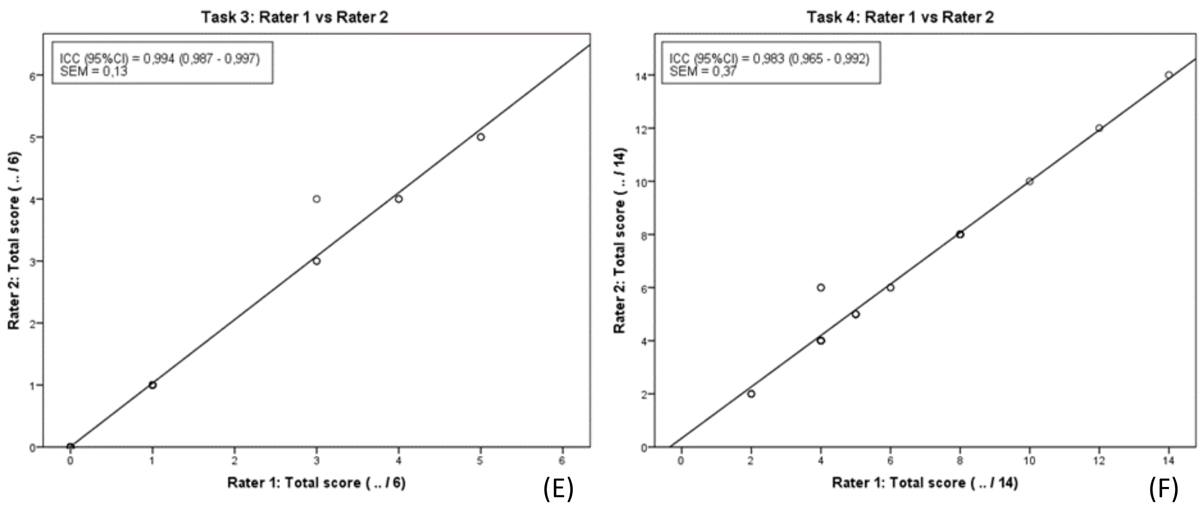
(A) (B)

Task 2 (up)		Rater 2 No pass	Rater 2 Pass	Total
Rater 1	No pass	7	0	7
	Pass	0	23	23
	Total	7	23	30

(C)

Task 2 (down)		Rater 2 No pass	Rater 2 Pass	Total
Rater 1	No pass	3	1	4
	Pass	0	26	26
	Total	3	27	30

(D)



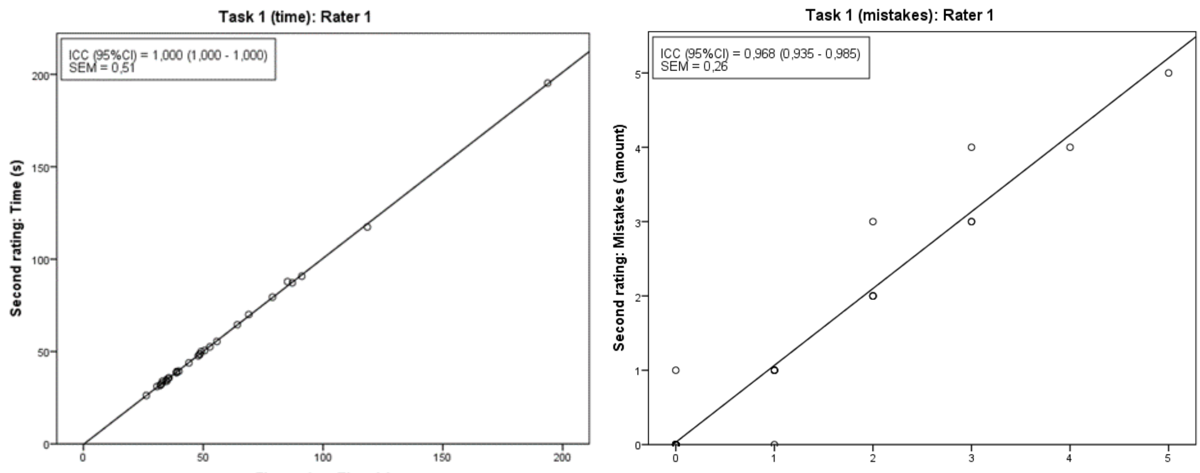
(E) (F)

Figure 2 Inter rater reliability of Rater 1 vs Rater 2, for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.

Table 4 Intra-rater reliability

Intra-rater reliability		Rater 1			Rater 2			Rater 3		
	N	ICC (95% CI)	SEM	κ	ICC (95% CI)	SEM	Κ	ICC (95% CI)	SEM	κ
Task 1										
Time	30	1,000 (1,000 – 1,000)	0,51		1,000 (1,000- 1,000)	0,30		0,998 (0,995 – 0,999)	1,68	
Mistakes	30	0,968 (0,935 – 0,985)	0,26		0,983 (0,965 – 0,992)	0,18		0,966 (0,930 – 0,983)	0,22	
Task 2										
Reaches the top	30			1,000			1,000			0,911[a]
Down with slalom	30			0,839[a]			1,000			0,870[a]
Task 3										
Total score	30	0,988 (0,975 – 0,994)	0,18		1,000[b]	– [b]		0,995 (0,990 – 0,998)	0,13	
Task 4										
Total score	30	0,977 (0,954 – 0,989)	0,43		1,000[b]	– [b]		0,977 (0,953 – 0,989)	0,43	

[a] 1 participant was scored different [b] No 95% CI is available because of perfect agreement, therefore the SEM can't be calculated either; ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement; κ = kappa



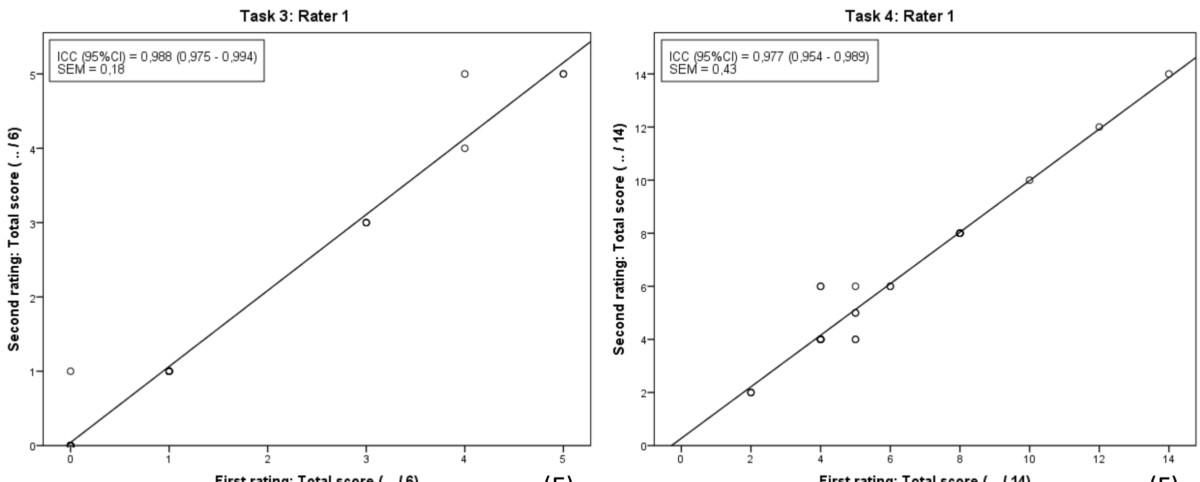
(A) (B)

Task 2 (up): Rater 1		Second Rating		
		No pass	Pass	Total
First Rating	No pass	7	0	7
	Pass	0	23	23
	Total	7	23	30

(C)

Task 2 (down): Rater 1		Second Rating		
		No pass	Pass	Total
First Rating	No pass	3	1	4
	Pass	0	26	26
	Total	3	27	30

(D)



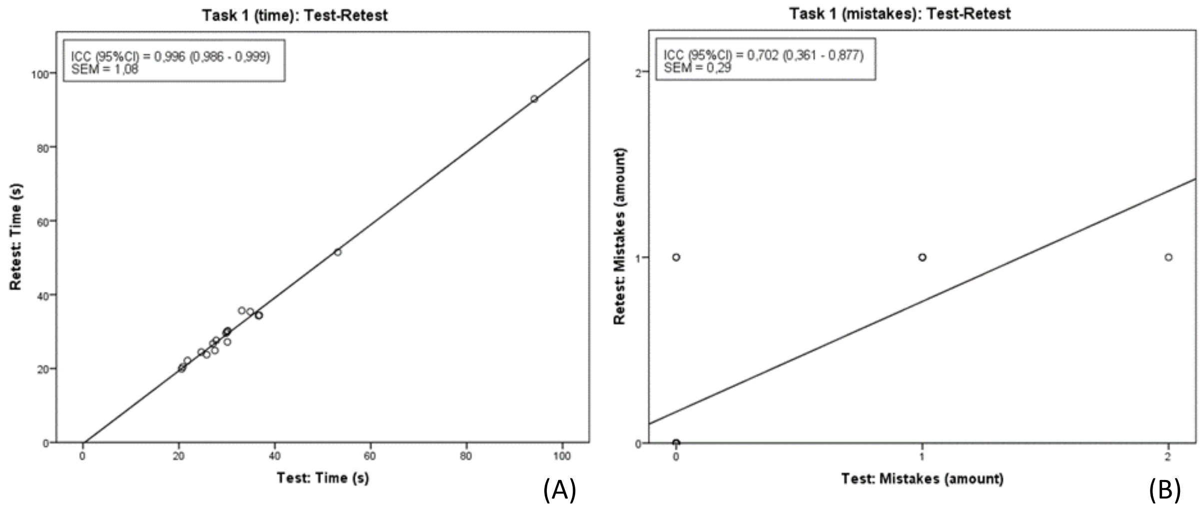
(E) (F)

Figure 3 Intra rater reliability of Rater 1, for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.

Table 5 Test-retest reliability

Test-retest reliability	N	ICC (95% CI)	SEM	κ
Task 1				
Time	18	0,996 (0,986 – 0,999)	1,08	
Mistakes	18	0,702 (0,361 – 0,877)	0,29	
Task 2				
Reaches the top	18			1,000
Down with slalom	18			1,000
Task 3				
Total score	18	1,000[a]	– [a]	
Task 4				
Kerb total score	18	0,992 (0,979 – 0,997)	0,33	

[a] No 95% CI is available because of perfect agreement, therefore the SEM can't be calculated either; ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement; κ = kappa



Task 2 (up)

		Retest: No pass	Pass	Total
Test	No pass	2	0	2
	Pass	0	16	16
	Total	2	16	18

(C)

Task 2 (down)

		Retest: No pass	Pass	Total
Test	No pass	2	0	2
	Pass	0	16	16
	Total	2	16	18

(D)

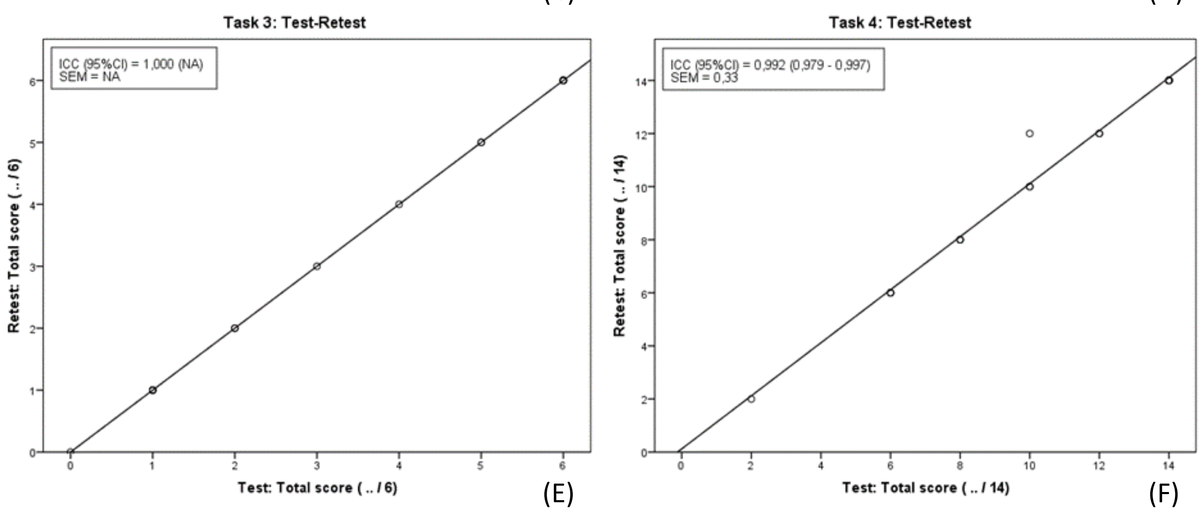


Figure 4 Test-retest reliability for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.

impact on the reliability. With a difference of three participants dropping the kappa to a moderate score of 0,609. Task 2 does show perfect kappa values for the test-retest reliability and strong to almost perfect for the intra-rater reliability. Therefore the task might need more explanation in the protocol, but interpretation of the current description seems to be clear even when scoring again after two weeks. It is also important to note that the reliability scores are calculated for these specific raters and are not directly generalizable to a broader population of raters. This is due to the fact that the raters were selected for their knowledge about wheelchair skills.

The DWMT is an assessment on which the first try counts, where the participant was explained and showed how to perform the task, but did not get a second try. This is unlike other wheelchair skill tests, where a participant gets either a time limit to perform the task correctly (31) or a second try when the first attempt failed (33). This might have influenced the results of the participant scores, by giving them a lower score than they might have been able to get. For determining the reliability this would not have been of any influence, since the raters still had to score if someone was capable of doing it or not. The second try was not used, because it was argued that a participant is skilful when the first try is successful. Another reason for not giving second tries was to reduce the total time necessary to complete the DWMT.

The test-retest reliability scores need to be interpreted with some caution, since the criteria for interpretation have not been met (41, 42). The sample size was lower than needed for interpretation, although with sufficient power according to the sample calculations (38). The values found are mostly excellent or almost perfect, from which it could be assumed that a sufficient sample size will result in at least good or strong reliability values.

A strength of the study is the diversity in the diagnosis of the participants. It is common to test a wheelchair skills test on participants with a spinal cord injury (29-32), whereas the current study included participants with a broad range of neurological disorders. The DWMT itself has a set-up that might also be usable for other diagnosis, although this needs to be tested. An example is people with stroke, who form a big part of the people with a neurological disorder (43), who have not yet been included. The reason for this is that the current protocol is not suitable for people who use one or both of their feet as well as their arms for wheelchair propulsion.

The DWMT set-up was done in such a way that distances for the manoeuvrability and precision tasks were based on real life distances, such as the width of a door (40). Although this was not possible in all of the situations, the slope used in the DWMT has a gradient of 1/10.3 instead of the Swedish standard of 1/12 (44). Though it is not likely for this difference in gradients to have an effect on the reliability of the test, rather on the performance of the participants.

The recruiting of the participants was done continuously and possible participants declined to participate in the study. Sometimes participants agreed on participation but did not show up, this was mainly at the test-retest. Because of the continuous recruitment this did not affect the final sample size, it might however have given a selection bias in the test. There are less people with MS who participated in the test-retest reliability compared to the inter- and intra-rater reliability sample. Since fatigue is a common symptom for people with MS (45) they might have scored differently on the test-retest if they did not have enough time to rest in between. It might be the case that the test-retest reliability for people who experience fatigue is not as good as the sample used in this study.

While this study determined the inter- intra- and test-retest reliability, and the validity relies on the adaptation of the QEWS, future research is needed to complete the psychometric properties of the DWMT. The sensitivity and specificity have not yet been determined, because this was outside of the scope of this study, although possible with the current data. The minimal clinical important difference needs to be determined if the DWMT will be used in effectiveness studies in the future.

Conclusion

The DWMT has good to excellent inter- intra-rater and test-retest reliability when sufficient sample sizes are used. The reliability was tested in a clinical situation which makes it suitable for further use in clinical practice to see what skill level a person has. Further research can be conducted using the DWMT, especially after determining the remaining psychometric properties.

References

1. Kirby RL, Mifflen NL, Thompson KJ, MacLeod DA. The manual wheelchair-handling skills of caregivers and the effect of training. Arch Phys Med Rehabil. 2004;85(12):2011-9.
2. World Health Organization. Fact sheet on wheelchairs. WHO; 2010.
3. Hjälpmedelsinstitutet. Förflyttnings och kognitionshjälpmedel - Metodbeskrivning statistik.: Hjälpmedelsinstitutet; 2014.
4. Flemmer CL, Flemmer RC. A review of manual wheelchairs. Disabil Rehabil Assist Technol. 2016;11(3):177-87.
5. van der Woude LH, Groot S. Wheelchair propulsion: a straining form of ambulation. Indian J Med res. 2005;121(121):719-22.
6. Chow JW, Levy CE. Wheelchair propulsion biomechanics and wheelers' quality of life: an exploratory review. Disabil Rehabil Assist Technol. 2011;6(5):365-77.
7. Alm M, Saraste H, Norrbrink C. Shoulder pain in persons with thoracic spinal cord injury: prevalence and characteristics. Journal of rehabilitation medicine. 2008;40(4):277-83.
8. Brose SW, Boninger ML, Fullerton B, McCann T, Collinger JL, Impink BG, et al. Shoulder ultrasound abnormalities, physical examination findings, and pain in manual wheelchair users with spinal cord injury. Arch Phys Med Rehabil. 2008;89(11):2086-93.
9. Curtis KA, Drysdale GA, Lanza D, Kolber M, Vitolo RS, West R. Shoulder pain in wheelchair users with tetraplegia and paraplegia. Archives of Physical Medicine and Rehabilitation. 1999;80(4):453-7.
10. Dyson-Hudson TA, Kirshblum SC. Shoulder pain in chronic spinal cord injury, part I: Epidemiology, etiology, and pathomechanics. Journal of Spinal Cord Medicine. 2004;27(1):4-17.
11. van Drongelen S, de Groot S, Veeger HEJ, Angenot ELD, Dallmeijer AJ, Post MWM, et al. Upper extremity musculoskeletal pain during and after rehabilitation in wheelchair-using persons with a spinal cord injury. Spinal Cord. 2006;44(3):152-9.
12. Boninger ML, Cooper RA, Robertson RN, Shimada SD. Three-dimensional pushrim forces during two speeds of wheelchair propulsion. American Journal of Physical Medicine & Rehabilitation. 1997;76(5):420-6.
13. Davidoff G, Werner R, Waring W. Compressive mononeuropathies of the upper extremity in chronic paraplegia. Paraplegia. 1991;29(1):17-24.
14. Pentland WE, Twomey LT. Upper-limb function in persons with long-term paraplegia and implications for independence. Paraplegia. 1994;32(4):211-8.
15. Sie IH, Waters RL, Adkins RH, Gellman H. Upper extremity pain in the postrehabilitation spinal-cord injured patient. Archives of Physical Medicine and Rehabilitation. 1992;73(1):44-8.
16. Meyers AR, Anderson JJ, Miller DR, Shipp K, Hoenig H. Barriers, facilitators, and access for wheelchair users: substantive and methodologic lessons from a pilot study of environmental effects. Social science & medicine (1982). 2002;55(8):1435-46.
17. Patel RM, Gelber JD, Schickendantz MS. The Weight-Bearing Shoulder. The Journal of the American Academy of Orthopaedic Surgeons. 2018;26(1):3-13.
18. Liu M, Mineo K, Hanayama K, Fujiwara T, Chino N. Practical problems and management of seating through the clinical stages of Duchenne's muscular dystrophy. Arch Phys Med Rehabil. 2003;84(6):818-24.
19. White HA, Lee Kirby R. Folding and unfolding manual wheelchairs: an ergonomic evaluation of health-care workers. Applied ergonomics. 2003;34(6):571-9.
20. Boninger ML, Towers JD, Cooper RA, Dicianno BE, Munin MC. Shoulder imaging abnormalities in individuals with paraplegia. Journal of rehabilitation research and development. 2001;38(4):401-8.
21. Gutierrez DD, Thompson L, Kemp B, Mulroy SJ, Physical Therapy Clinical Research N, Rehabilitation R, et al. The relationship of shoulder pain intensity to quality of life, physical activity, and community participation in persons with paraplegia. The journal of spinal cord medicine. 2007;30(3):251-5.

22. Sawatzky B, DiGiovine C, Berner T, Roesler T, Katte L. The need for updated clinical practice guidelines for preservation of upper extremities in manual wheelchair users a position paper. American Journal of Physical Medicine & Rehabilitation. 2015;94(4):313-24.
23. Kendall MB, Ungerer G, Dorsett P. Bridging the gap: transitional rehabilitation services for people with spinal cord injury. Disability and rehabilitation. 2003;25(17):1008-15.
24. McNevin NH, Wulf G, Carlson C. Effects of attentional focus, self-control, and dyad training on motor learning: implications for physical rehabilitation. Physical therapy. 2000;80(4):373-85.
25. Routhier F, Vincent C, Desrosiers J, Nadeau S. Mobility of wheelchair users: a proposed performance assessment framework. Disability and rehabilitation. 2003;25(1):19-34.
26. World Health Organization. Wheelchair service training package trainer's manual. Malta: WHO; 2012.
27. Norsten Å. Drivkraft - körergonomi, rullstolteknik & metodik. Västerås: Westerås Media Produktion AB; 2001.
28. Zanca JM, Natale A, Labarbera J, Schroeder ST, Gassaway J, Backus D. Group physical therapy during inpatient rehabilitation for acute spinal cord injury: findings from the SCIRehab Study. Physical therapy. 2011;91(12):1877-91.
29. Fliess-Douer O, Vanlandewijck YC, Lubel Manor G, Van Der Woude LH. A systematic review of wheelchair skills tests for manual wheelchair users with a spinal cord injury: towards a standardized outcome measure. Clin Rehabil. 2010;24(10):867-86.
30. Kirby RL, Worobey LA, Cowan R, Pedersen JP, Heinemann AW, Dyson-Hudson TA, et al. Wheelchair skills capacity and performance of manual wheelchair users with spinal cord injury Archives of Physical Medicine and Rehabilitation. 2016;97(10):1761-9.
31. Cowan RE, Nash MS, de Groot S, van der Woude LH. Adapted manual wheelchair circuit: test-retest reliability and discriminative validity in persons with spinal cord injury. Archives of Physical Medicine and Rehabilitation. 2011;92(8):1270-80.
32. Gollan EJ, Harvey LA, Simmons J, Adams R, McPhail SM. Development, reliability and validity of the Queensland evaluation of wheelchair skills (QEWS). Spinal Cord. 2015;53(10):743-9.
33. Kirby RL, Smith C, Parker K, MacLeod DA, McAllister M. Wheelchair Skills Test (WST) Version 4.2 Manual. Halifax, Canada: Dalhousie University; 2013.
34. British Standards Institution. BS EN 1340:2003 Concrete kerb units - Requirements and test methods. London, UK; 2003.
35. Cowan RE, Callahan MK, Nash MS. The 6-min push test is reliable and predicts low fitness in spinal cord injury. Medicine and science in sports and exercise. 2012;44(10):1993-2000.
36. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2002;11(3):193-205.
37. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. Statistical Methods in Medical Research. 2004;13(4):251-71.
38. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Statistics in Medicine. 1998;17:101-10.
39. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Archives of Orofacial Sciences. 2017;12(1):1-11.
40. SiS. Dörrar - Väggöppningar för dörrar - Mått. Sweden: SiS; 1986.
41. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of chiropractic medicine. 2016;15(2):155-63.
42. McHugh ML. Interrater reliability: the kappa statistic. Biochemia medica. 2012;22(3):276-82.
43. Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. Lancet (London, England). 2008;371(9624):1612-23.
44. Boverket. Tillgänglighet - Ramper. PBL kunskapsbanken - en handbok om plan- och bygglagen. Sweden: Boverket; 2019.
45. Loy BD, Taylor RL, Fling BW, Horak FB. Relationship between perceived fatigue and performance fatigability in people with multiple sclerosis: A systematic review and meta-analysis. J Psychosom Res. 2017;100:1-7.

Appendix 1 Protocol: How to score, measure & instruct

Drivkraft Wheelchair Manoeuvre Test

How to score, measure & instruct

Name: ______________________

Date: ________________________

Assessor: ____________________

Before starting the measurement:

- Check if the participant has signed the informed consent.
- Check if the participant understands what is going to happen.
- Check if all the details are filled in on the reporting form.
- Check if the person has anti-tippers and that they are in use. Otherwise a person is needed to secure the safety.

In general:

- Each participant is entitled to a full explanation of what is expected of him/her.
    - This explanation can include an example performance by the assessor.
- Questions for clarification may be asked and answered in addition to the standard instructions.
- Every measurement has 1 attempt, no retakes are allowed.
    - Exceptions can be made if it is due to problems with the equipment, mechanics of the wheelchair or a clear misunderstanding of the task. A mark will need to be made to state the reason for giving the second attempt.
- A countdown is performed before every attempt to let the participant know when he/she can start.

For the assessors:

- It's important to not say out loud what is being scored. This could influence the participant, but moreover it will influence the other raters who will have to judge from the videorecording.
- When rating from the video you are not allowed to use the pause or playback function.

Task 1: Fundamental manoeuvring – *Precision*

Time: _______ sec
Mistakes: _______ (amount)

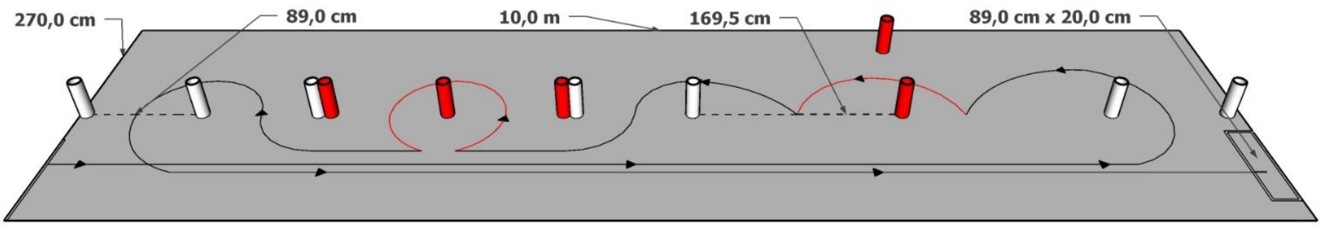
Instructions:

All the white cylinders are supposed to be passed forward and the red cylinders you have to go backwards.
You start at the line here, with the casters behind the line, and you stop with the casters in the box over there and the aim is not to hit the line.
I will show you first, then you can try and then it will be on time.
So I count down 3 2 1 go
/explain what you are doing while showing/
/let the participant try and explain what he/she is supposed to do/
So now you will do it once more but on time.

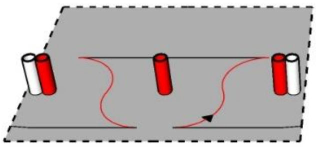
Scoring:

- The time starts after a countdown of 3-2-1-go, the time finishes when the participants stops the wheelchair at the finish box.
- The starting position is behind the line, the casters can not touch the line.
- If the participant doesn't stop between the two lines it counts as a mistake.
- Passing the box is therefore a mistake, even when going backwards after breaking behind the box. Time stops when the wheelchair stops and it counts as a mistake if it's not inbetween the finish lines.
- Touching a cylinder is a mistake, the cylinder does NOT have to fall completely for it to count as a mistake.
- In case the cylinder falls, the assessor or an assistant can quickly pick it up to prevent it from being an obstacle later on.
- The double red gates can be passed two ways, both routes are shown in the figure below. (red lines = backwards, black lines = forward)
- If the participant goes the wrong way in the slalom, he/she will be corrected by the assessor to go the other way. It does however count as a mistake.

Set up:



Or



Task 2: Fundamental manoeuvring – *Slope of 1/10,3*

a. Reaches the top: Yes / No

Instructions:

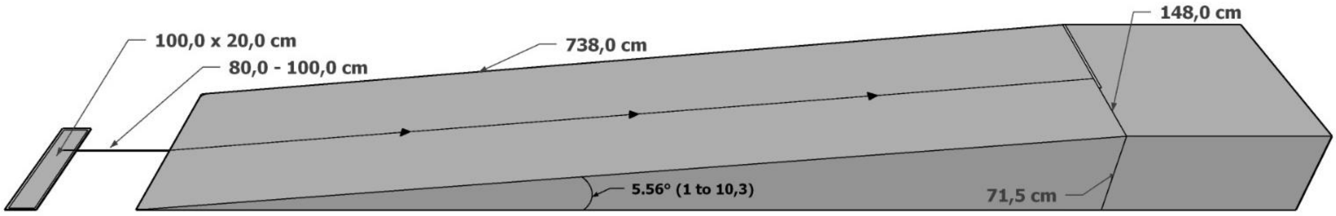
The aim is to go up the slope, and you stop when you are at the top.
You start with the casters in the box, which you can do now.

Scoring:

- It only counts as going up if the participant reaches the platform on the top.
- If the participant doesn't reach the top it is important that he comes down safely. To make sure of this the assessor or assistant should be walking behind him/her.

! The participant is during the whole time in charge of his/her feeling of safety and can stop at any time, then the assessor or assistant will take over to get the participant down safely.

Set up:



** The pictures of the slopes don't show it, but there are walls on each side of the slope, so there is no risk of falling of on the sides.*

---

b. Goes down the slope with slalom: Yes / No
Mistakes: _______ (amount)

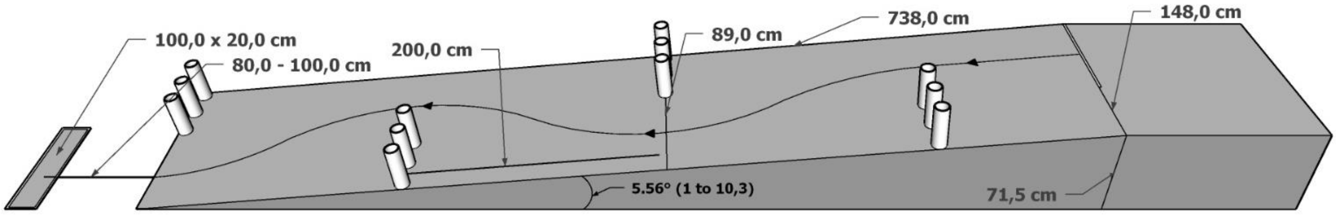
Instructions:

Now we go down the slope with the slalom. You can't touch the walls or the cylinders and the goal is to stop with the casters inside the box in a controlled way. Lean backwards while going down and keep sliding/breaking. You are allowed to push on the flat surface to reach the box.

Scoring:

- Touching a cylinder or the wall is a mistake, the cylinder does NOT have to fall completely for it to count as a mistake.
- Going down needs to be done with a sliding/breaking grip, pushes for steering are allowed.
- If the participant doesn't stop between the two lines it counts as a mistake.
- It is allowed to push on the flat surface to reach the two lines. The participant doesn't need to do this only by having the remaining speed of the slope.
- Mistakes result in a result of 'No pass' on the slalom.

Set up:



Task 3: Control of centre of gravity – *Rear wheel balancing, push and sliding/breaking*

- Every part earns a point when passed which will be added to each other at the end.
- In all of the situations the casters are not allowed to touch the ground unless specified otherwise.
- In all of the situations the anti-tippers are not allowed to touch te ground.

- Push 1: Rear wheel balance for 20 seconds

Instructions:

you will balance for 20 seconds on your rear wheels. I will say when the time has passed so you can put them down again. You will need to stay inside the square but you are allowed to move in the square as long as your rear wheels don't go out.
Start in the centre of the square and then I will start the countdown for you to start the balancing.

Scoring:

- Time starts when the casters leave the ground.
- After 20 seconds the participant can stop, there is no need to continue for a longer amount of time.
- The participant is allowed to move around on the spot as long as the rear wheels are not leaving the $1m^2$ square.
- It is not allowed for the rear wheels to touch the line of the square.
- If the participant is not able to complete this task, the rest of Task 3 will be scored as a no pass. This is due to safety reassurance and the skills only get harder from here.

- Push 2. Turn 360 degrees (in both directions)

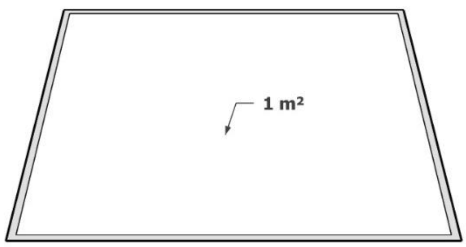
Instructions:

For this part you have to stay inside the square with your rear wheels. You go on your rear wheels and you turn either left or right, after that I count 3 2 1 and you can turn the other way. Then I will count again 3 2 1, and you can put the casters down.

Scoring:

- The participant can use as many pushes as he/she wants.
- The rear wheels can not leave the $1m^2$ square.
- It is not allowed for the rear wheels to touch the line of the square.
- The casters are allowed to hover over the square.
- The participant needs to stay still for 3 seconds after the first 360 before turning the other way.
- The participant needs to stay still for 3 seconds when finished with the second 360.

Set up for both Push 1 and Push 2:



- Push 3. High / Low balance position 10 cm

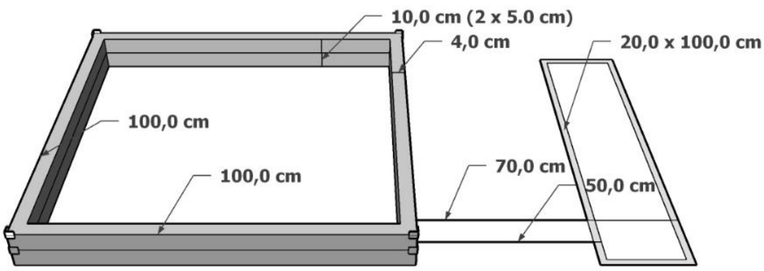
Instructions:

For this part you start with the casters in the box and you start to balance on the rear wheels. You drive forward on the rear wheels to the frame and you put the casters down on the other side of the frame. Then you go back in the rear wheel balance and you drive backwards. To show me that you can keep control, I will count 3 2 1 and then you can put the wheels down in the box.
Note: Check if the participant has anti-tippers and if they will allow a sufficiently high balance position to pass this task. Change the settings of the anti-tippers and provide safety in case they were not high enough.

Scoring:

- Casters are supposed to touch the ground after moving over the obstacle.
- Casters shouldn't be resting on the obstacle, touching/tapping on the obstacle is allowed.
- Rear wheels can be touching the obstacle.
- When going backwards a 3 second stop on the rear wheels is needed to show control before putting the casters down.
- If the participant notices he/she went too far backwards he/she can correct it before taking the 3 second stop. This would still count as being in control and being precise.

Set up:



- Sliding/breaking 1. Slope on the rear wheels

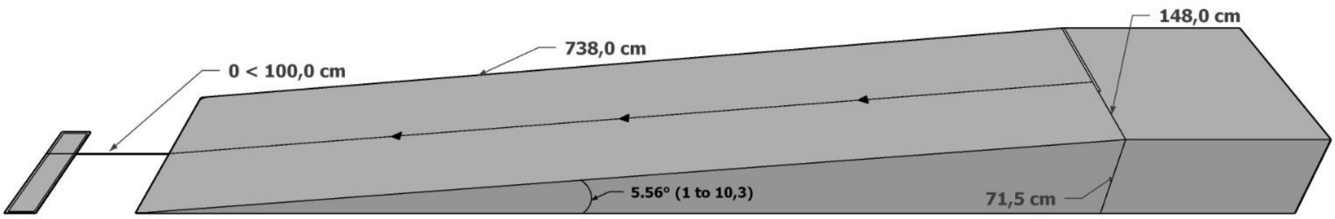
Instructions:

Now we will go down the slope on the rear wheels, it's important to keep sliding/breaking and not push. When you come down it's important to stop on the rear wheels while maintaining control, this has to be before the box. You cannot touch the line of the box with the rear wheels. I will count down again 3 2 1 and then you can put the wheelchair down.

Scoring:

- Goes down with a sliding/breaking grip, small pushes to control the center of gravity are allowed.
- When at the bottom of the slope a 3 second stop on the rear wheels has to be performed to show control.
- The participant should stop before the last line of the box, which is the 1 meter line to show they have control over their speed. They can't touch this line with their rear wheels.
- In case of a no pass, Sliding/breaking 2 and 3 do not have to be performed.

Set up:



- Sliding/breaking 2. Slope on the rear wheels with slalom

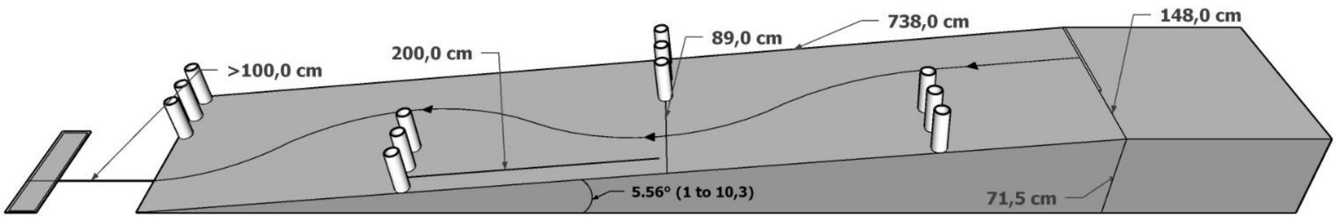
Instructions:

During this test you will again go down on the rear wheels while sliding/breaking. You can't push during this test. The goal is to pass the box with the rear wheels without extra pushes and then you stop in a controlled way. I will count down 3 2 1 and you can put the casters down.

Scoring:

- When the participant has traveled the 1 meter he/she should remain in the rear wheel balance for 3 seconds before touching the ground to show full control.
- It's not allowed to push in the 1 meter that is on flat surface, this distance has to be covered by the speed of the slope and the control of sliding/breaking.
- The 1 meter line needs to be passed with the rear wheels. Which means no touching of the line with the rear wheels.
- When the participant touches one of the cylinders or the wall it's a fail.

Set up:



- Sliding/breaking 3. Flat surface distance of 2 meters on rear wheels

Instructions:

You start behind the line with the casters here and you gather speed. Before the line you lift your casters up and go the whole way on your rear wheels while sliding/breaking, until you pass the other line with your rear wheels. After passing the line you stop while staying in control and I will count down 3 2 1 and you can put the casters down.

Scoring:

- If the participant doesn't lift the casters before the line it's a fail.
- The finish line needs to be passed with the rear wheels, if the participant drops the casters before the finish line it's a fail.
- If the participant passes the 2 meters, but can't control the wheelchair for 3 seconds it's a fail.
- If the participant starts by lifting the casters before the line and puts them down after the finish line, which will be more than 2 meters, it's a pass.
- No pushing is allowed while on the rear wheels.
- A maximum of 3 meters is allowed to gain speed.

Set up:

Total: ___/6

Task 4: Obstacle technique – Going up and down a kerb

Kerb height	Up	Down
1.0 cm	o	o
2.5 cm	o	o
5.0 cm	o	o
7.5 cm	o	o
10.0 cm	o	o
12.5 cm	o	o
15.0 cm	o	o

Total: ___/14

Instructions:

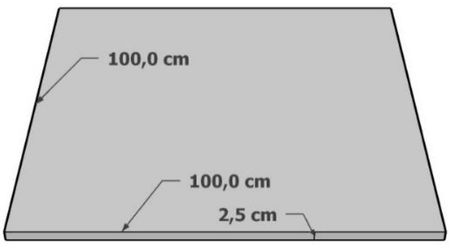
Now we will go up a kerb. We start with 1cm and go up to 2.5, 5, 7.5, 10, 12.5 and 15cm as long as you keep getting up and feel safe to try. You start here in the box and you can gather speed to go up the kerb. You can also drive towards the kerb and stop to get up, both ways are allowed. When you go down you have to land on your rear wheels.
Note: When someone doesn't get up explain he can try one more time.
Note: Check if the participant has anti-tippers and if they will allow a sufficiently high balance position to pass this task. Change the settings of the anti-tippers and provide safety in case they were not high enough.

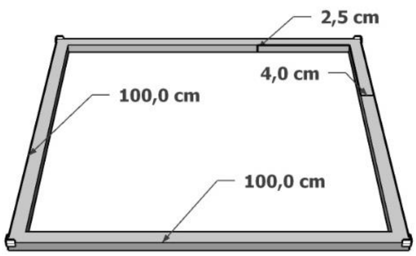
Scoring:

- A participant can continue to the next height when the previous height was succesful on the first attempt. i.e. a participant goes up the 1.0cm kerb on the first go, goes up the 2.5cm kerb on the first go, but fails the first attempt on the 5.0cm. In this case the participant can try one more time the 5.0cm if he/she manages to get up this time the final height will be 5.0cm and the test stops. If he/she doesn't manage on the second attempt the final height is 2.5cm.
- Going down is considered to be safe when the participant goes down on the rear wheels or lands with the casters and rear wheels rather simultaneously.
    - The rear wheels shouldn't be on top of the kerb while the casters are touching the ground.
- If it's decided that the participant doesn't go down in a safe way it will be scored as a fail. A comment will be made by the assessor to state it wasn't safe. Going up the kerb to the next stage is allowed, but going down will be done with support by lowering the kerb back to 0. The participant can continue trying going further up, but can no longer go down on his/her own.
- A total of 3 meters can be used to gain speed to get up the kerb, this depends on personal preferences on what distance will be used.
- If the participant estimated it wrong and stops in front of the kerb it's a fail. He/she can continue to the next level to see if he/she can pass it and then continue from there.
    - Depending on the technique a person can stop in front of the kerb, put the casters on the kerb and create a momentum from there to get up. This will involve moving backwards while the casters remain on the kerb and a strong push forward to get up completely. This can also be done just once, otherwise it's a fail.

Set up:

A combination of kerb heights and frames are used to create the heights that are tested. See list of equipment for an overview of the sizes used.





Appendix 2 Protocol: How to film

Drivkraft Wheelchair Manoeuvre Test

How to film

For the filming of the test it's important to keep as close to Åke as possible, to keep his point of view. If another angle is created the second and third rater might be seeing more, or less, then the first rater. It is therefore also important to focus the filming on the parts that are important. If there is a focus on the wheels, either the rear wheels or the casters, that need to start or end at a line it should be visible on the video as well. The video is therefore mobile, and will move along the test.
The entire test will be filmed, without stopping. This is done so that the total duration of the test will be known, this can be used to determine the average time to take the entire test. Before sending it to the third rater it will be edited to only show the test components to save time.

Task 1:
Walk with the participant over the 10m length of the course, make sure that the participant is visible at all times and especially the wheels need to be clear on video. When at the finish box make sure that the whole box is visible on camera to see if the participant stops inside the box with the casters.

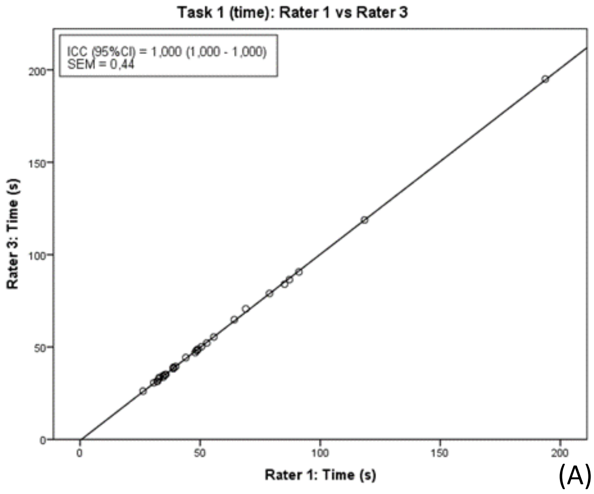
Task 2:
Going up the slope the camera will be positioned at the bottom of the slope, no movements need to be made to see if the participant reaches the top.
Going down the slope with the slalom the camera will start at the bottom of the slope, but will move to the side once the participant gets down. This is to show if the participant stops with the casters inside the box. It's therefore important that the entire box is visible on the video.

Task 3:
The first three components are easy to film because they require little space. Make sure the participant and the lines are visible on the video and no other movements need to be made.
For the fourth and fifth component, which are down the slope, the same applies as with task 2. The camera will start at the bottom of the slope, but will move to the side once the participant gets down. All lines need to be visible when the participant stops.

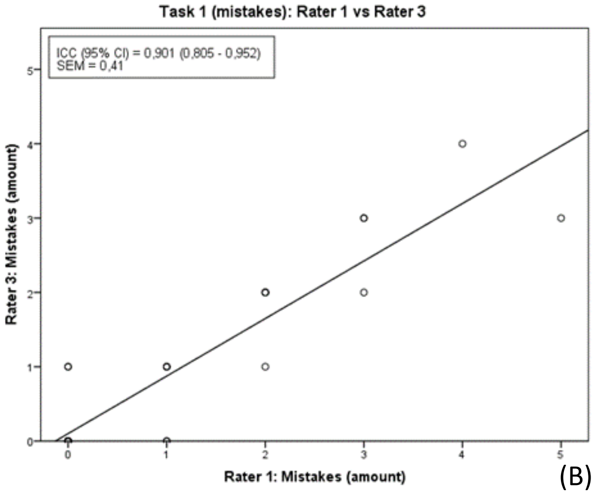
Task 4:
Not much space is needed, therefore there is no need to be moving around with the camera. The camera should follow the participant during the 3 meters heading towards the kerb and can be static if the whole kerb is visible so that both the going up and down will be visible.

Appendix 3 Inter-rater reliability figures

Rater 1 vs Rater 3



(A)



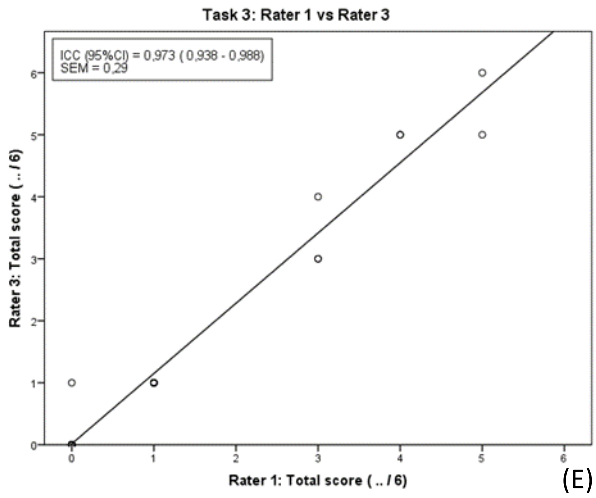
(B)

Task 2 (up)		Rater 3 No pass	Rater 3 Pass	Total
Rater 1	No pass	7	0	7
	Pass	1	22	23
	Total	8	22	30

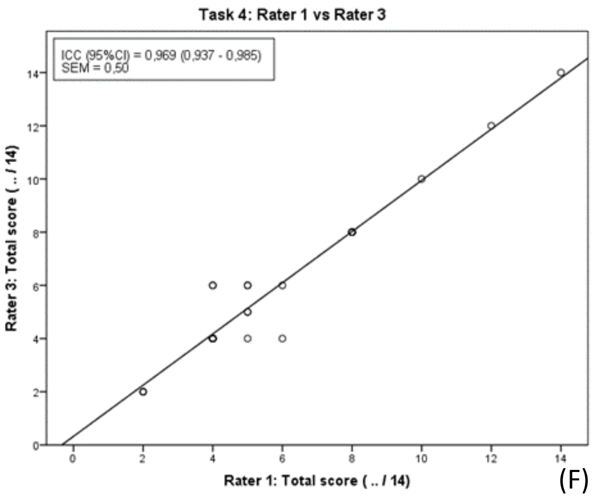
(C)

Task 2 (down)		Rater 3 No pass	Rater 3 Pass	Total
Rater 1	No pass	3	1	4
	Pass	2	24	26
	Total	5	25	30

(D)



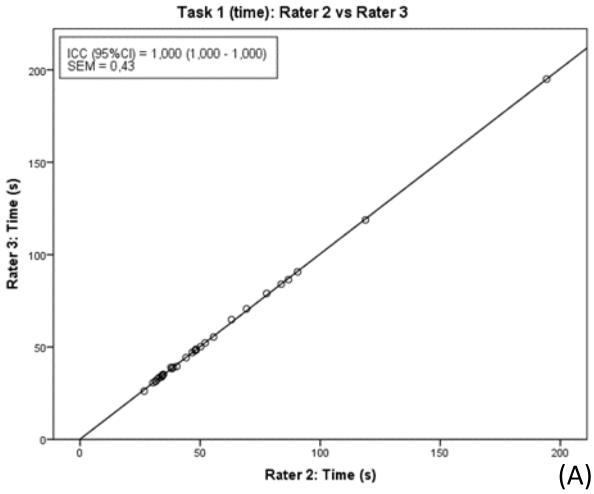
(E)



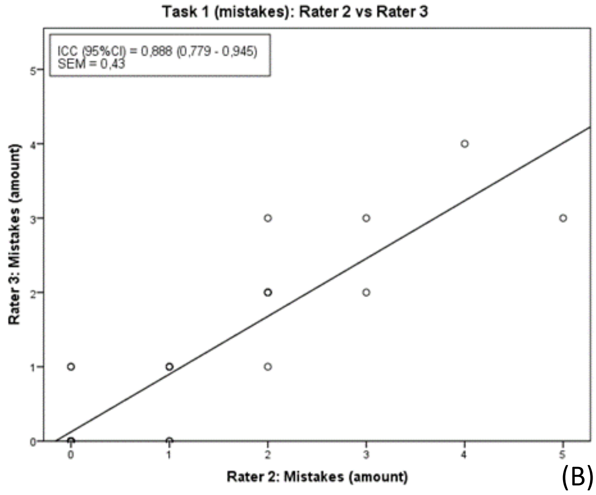
(F)

Figure A1: Inter rater reliability of Rater 1 vs Rater 3, for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.

<u>Rater 2 vs Rater 3</u>



(A)



(B)

Task 2 (up)

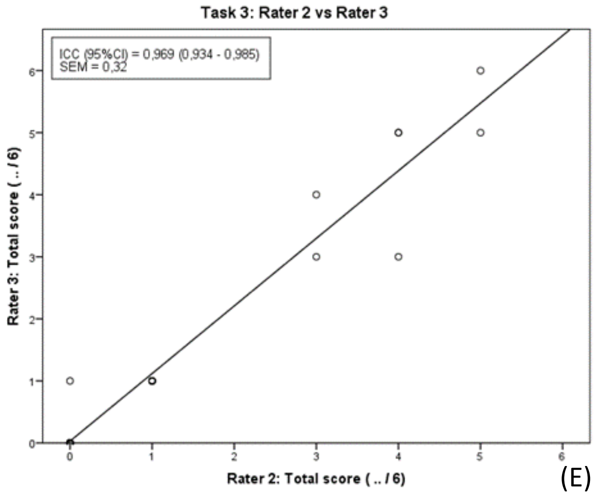
		Rater 3 No pass	Rater 3 Pass	Total
Rater 2	No pass	7	0	7
	Pass	1	22	23
	Total	8	22	30

(C)

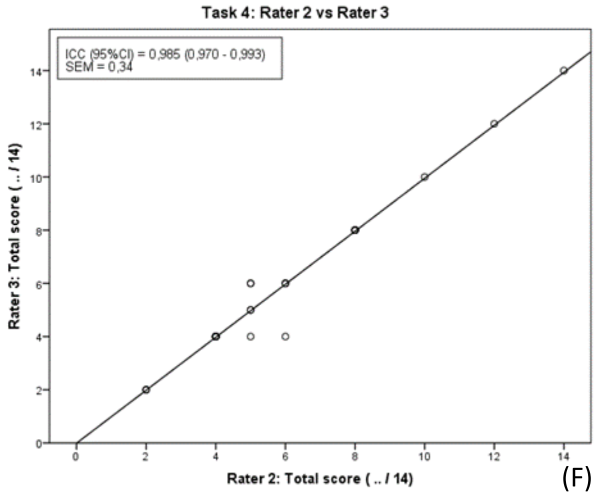
Task 2 (down)

		Rater 3 No pass	Rater 3 Pass	Total
Rater 2	No pass	3	0	3
	Pass	2	25	27
	Total	5	25	30

(D)



(E)

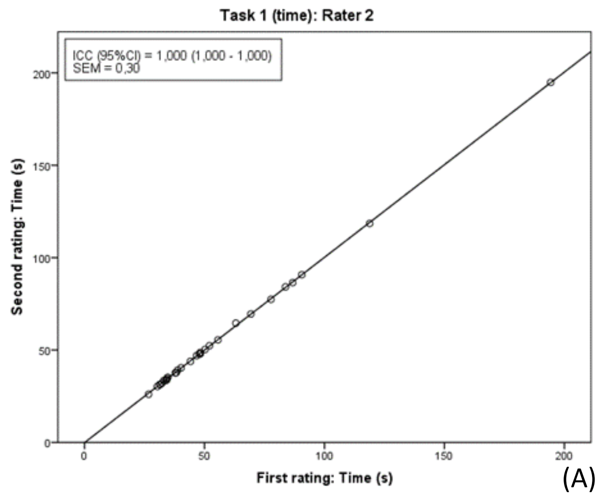


(F)

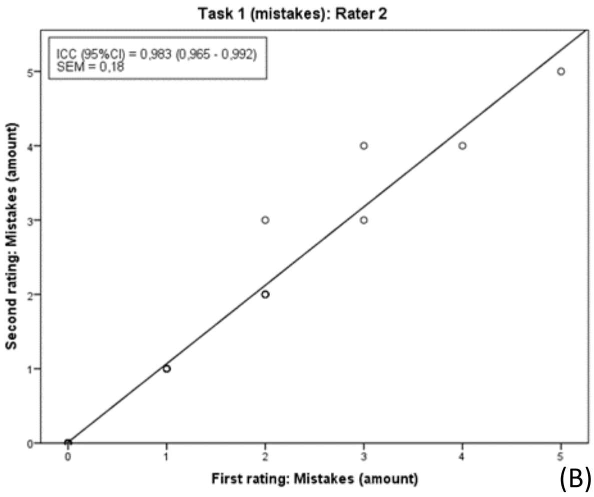
Figure A2: Inter rater reliability of Rater 2 vs Rater 3, for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.

Appendix 4 Intra-rater reliability figures

Rater 2



(A)



(B)

Task 2 (up): Rater 2

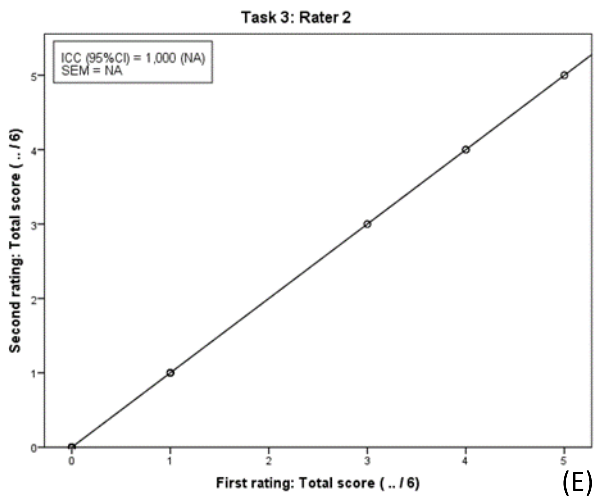
		Second Rating		
		No pass	Pass	Total
First Rating	No pass	7	0	7
	Pass	0	23	23
	Total	7	23	30

(C)

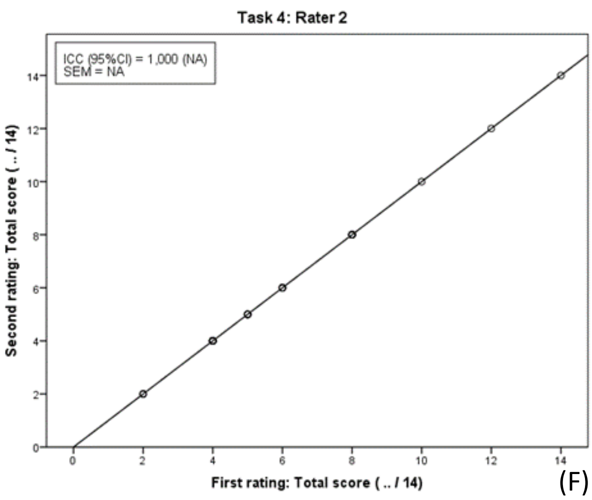
Task 2 (down): Rater 2

		Second Rating		
		No pass	Pass	Total
First Rating	No pass	3	0	3
	Pass	0	27	27
	Total	3	27	30

(D)



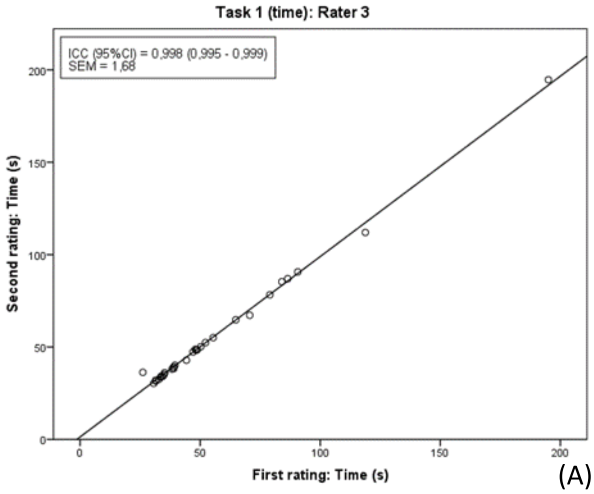
(E)



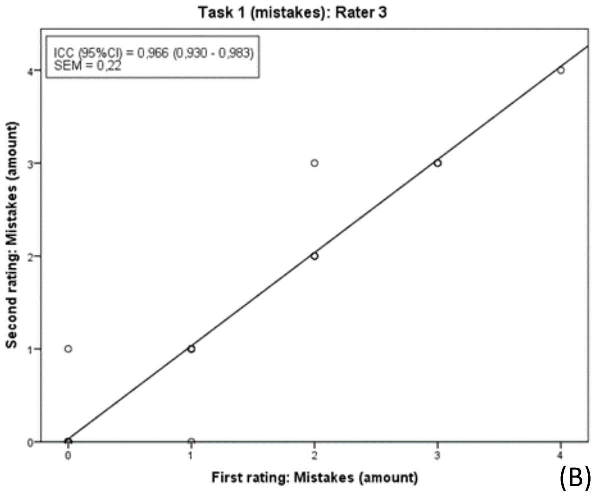
(F)

Figure A3: Intra rater reliability of Rater 2, for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.

Rater 3



(A)



(B)

Task 2 (up): Rater 3

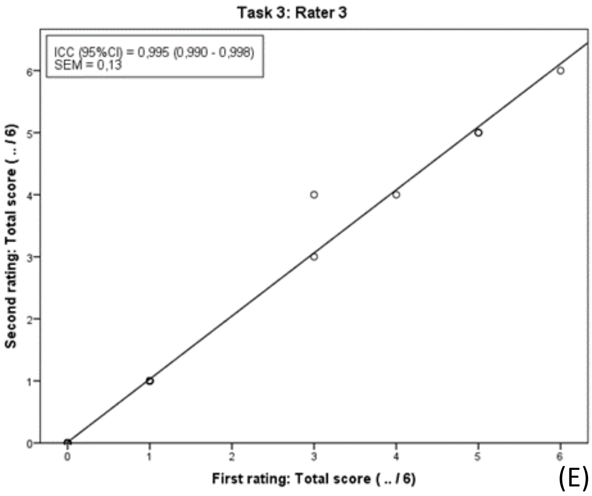
		Second Rating No	Yes	Total
First Rating	No	7	1	8
	Yes	0	22	22
	Total	7	23	30

(C)

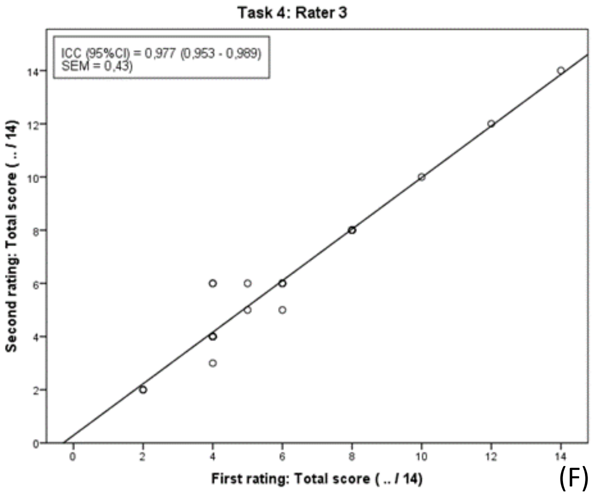
Task 2 (down): Rater 3

		Second Rating No	Yes	Total
First Rating	No	4	1	5
	Yes	0	25	25
	Total	4	26	30

(D)



(E)



(F)

Figure A4: Intra rater reliability of Rater 3, for each task. ICC = intraclass correlation coefficient; 95% CI = 95% confidence interval; SEM = standard error of measurement.