Reliability of the Woods and Teuber scale: Assessment of mirror movements in children and adolescents with unilateral cerebral palsy

VICTORIA A. MAGNE[1], LARS ADDE[2,3], BRIAN HOARE[4,5], KATRIJN KLINGELS[6,7], CRISTINA SIMON-MARTINEZ[6,8], LISA MAILLEUX[6], STIAN LYDERSEN[9], ANN-KRISTIN G. ELVRUM[2,3]

[1]Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway
[2]Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway
[3]Clinical Services, St. Olavs Hospital, Trondheim University Hospital, Norway
[4]La Trobe University, School of Occupational Therapy, Victoria, Australia
[5]Department of Paediatrics, Monash University, Clayton, Victoria, Australia
[6] KU Leuven – University of Leuven, Department of Rehabilitation Sciences, Leuven, Belgium
[7]Faculty of Rehabilitation Sciences, Rehabilitation Research Center (REVAL), Hasselt University, Diepenbeek, Belgium
[8]Institute of Information Systems, University of Applied Sciences Western Switzerland Valais, Sierre, Switzerland
[9]Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence to
Ann-Kristin G. Elvrum,
Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Post-box 8905,
NO-7491 Trondheim, Norway

E-mail: ann-kristin.elvrum@ntnu.no

https://orcid.org/0000-0001-9712-1447

Word count: 3044

Abstract

Aim: The aim of this study was to investigate inter- and intrarater reliability of the Woods and Teuber scale (W&T) to detect mirror movements (MMs) in children and adolescents with unilateral cerebral palsy (UCP).

Method: A convenience sample of children and adolescents with UCP (n=68; mean age 12 years (y) and 2 months (m), SD 3y6m; 31 boys) and Manual Ability Classification levels I-III was recruited from Norway, Australia and Belgium. Three therapists scored MMs according to the W&T scale from three video-recorded tasks, at two separate time points. A two-way mixed model regression was used to calculate intraclass correlation coefficients (ICCs) reflecting overall inter- and intrarater reliability. In addition, ICCs for each hand and task separately were calculated.

Results: The overall interrater reliability ICC was 0.90, and the corresponding intrarater reliability ICC was 0.92. The ICCs for each hand ranged from 0.86 to 0.92 and for each task from 0.63 to 0.89.

Interpretation: The W&T scale shows excellent reliability for scoring MMs in children and adolescents with UCP. The assessment is easy to administer with no need for specific equipment and scoring can be determined from short video-recordings making it a feasible instrument in research and clinical practice.

What this paper adds:

- The Woods and Teuber scale produces reliable measures of mirror movements.
- Scorings can reliably be determined from short videos recorded from above.
- The Woods and Teuber scale has good clinical utility.

Running foot: Reliability of Woods and Teuber scale

Mirror movements (MMs) are involuntary movements activated by voluntary movements of the contralateral limb.[1,2] They are more often observed in the upper limbs and intensity may increase due to growing task complexity or fatigue.[3] MMs are present in typically developing children, but decrease gradually between five and eight years of age and have usually disappeared by 10 years of age.[4]

In many children with unilateral cerebral palsy (UCP), MMs are more pronounced and often persist after the age of 10.[1,5,6] Two underlying mechanisms are postulated for the presence and persistence of MMs in these children.[2,7,8] First, a reorganization in the corticospinal system of the motor cortex can result in both hands being controlled by the less-affected brain hemisphere.[2,6,8] This is thought to cause MMs primarily in the more-affected hand.[6,8] Second, a dysfunction of the corpus callosum may result in bilateral cortical activation due to imbalance of interhemispheric inhibition.[7] This bilateral activation of the motor cortices may lead to motor overflow causing MMs in the less-affected hand when intentionally moving the more-affected hand.[3,7,8] Studies have demonstrated that MMs seem to be more prominent in the less-affected hand i.e. when moving the more-affected hand, and that MMs in this hand are more often related to impaired bimanual performance.[3,5,7,9-11] Results regarding the association between MMs in the more-affected hand and bimanual performance are more inconclusive.[3,5,9,12,13] MMs may assist in activities requiring symmetric hand use, but may cause difficulties performing activities in daily life where asymmetric hand use is required.[7] Moreover, it has been suggested that the response to different kinds of intensive therapy programs are influenced by the degree of MMs in children with UCP.[5,14-17] Therefore, reliable assessment of MMs in future studies of upper limb intervention may assist in better understanding their influence on outcomes.[5,14] The most commonly used instrument to detect the presence of MMs in children with UCP is the Woods and Teuber (W&T) scale.[1]

The W&T scale is an observation-based assessment with five distinct levels describing the degree of MMs.[1] The scale was developed in 1978 and measures MMs observed in the resting hand, while the other hand is moved intentionally performing three instructed tasks.[1] Originally, the W&T scale was administered using these tasks: (1) repetitive tapping of the index finger on the thumb, (2) alternating supination and pronation of the hand, and (3) repetitive alternate touching of each fingertip. Since then, variants of the tasks have been administered (i.e. fist opening and clenching, and finger tapping on the table). More recently, laboratory-based assessments, like the Windmill task,[9] and grip force devices[13,18] have also been used as alternative methods to measure MMs. These methods provide objective quantification of MMs but require computerized equipment that is not clinically available. The W&T scale does not require specific equipment and remains the primary clinical method to detect MMs.[19] It has been used in numerous studies to investigate the relationship between MMs and brain damage, motor impairments or bimanual performance in children with UCP.[5,8,13,11]

While the W&T scale has been used clinically and in research for more than four decades, information regarding the measurement properties of the scale remains largely unknown.[7] In 2016, Klingels et al.[7] reported high intra- and interrater reliability with intraclass correlation coefficients (ICCs) of >0.82. Furthermore, Riddell et al.[10] recently reported high interrater reliability (ICC: 0.87) for a modified W&T scale. However, the main purpose of these two studies were not to investigate reliability. The reliability analysis was performed on a subsample of only 20 children and included two raters in the study by Klingels et al.,[7] while Riddell et al. included three raters with only 18 participants.[10] Moreover, scarce information was provided regarding the procedures and sample characteristics.[7,10] Further investigation of the measurement properties of the W&T scale is therefore required to confirm these results.[20] The aim of this study was to investigate intra- and interrater reliability of the W&T scale, when assessing MMs in a large group of children and adolescents with UCP.

Method

This is a multicenter study with a cross-sectional design using data from prior studies performed in Norway, Australia and Belgium. All participants were recruited through convenience sampling. The Norwegian participants were recruited from the outpatient clinic at St. Olav University Hospital in Trondheim between 2011 and 2015, the participants from Australia were recruited at Monash Children´s Hospital in Melbourne between 2015 and 2018. Finally, the Belgian participants were recruited at the University Hospitals Leuven between 2014 and 2017. The participants were eligible to participate if they met the following criteria: diagnosis of spastic UCP, aged 5–20 years at the time of recruitment and able to sufficiently cooperate to complete the assessment. Participants were excluded if they had undergone upper limb surgery within 12 months of assessment or been injected with Botulinum toxin-A within three to six months of assessment.

Ethical approval was granted at all sites by the Regional Ethics Committee in Mid-Norway (REK 2011/1451 and 2015/1744), the Human Research Ethics Committee at Monash Children's Hospital, Victoria (HREC approval 12167B), and the Ethics Committee at the University Hospitals Leuven (S55555 and S56513). All participants agreed to participate in the study and provided written informed consent (including written consent from the parents of teenagers under the age of 16). In the current study, all data used were de-identified using an ID code for each participant.

Data collection

Background variables such as gender, age and side of hemiplegia were registered. In addition, the severity of hand function was described by means of the Manual Ability Classification System (MACS) levels and the Assisting Hand Assessment (AHA).[21,22]

To collect data on MMs, three repetitive tasks based on the W&T assessment were performed and videotaped for later scoring. For the Norwegian and Australian participants, the following tasks were performed: (I) fist opening and clenching; (II) repetitive tapping of the index finger on the thumb; and (III) simultaneously tapping fingers on the table. The Belgian participants performed (I) fist opening and clenching; (II) opposition of four fingers sequentially to the thumb; and (III) sequentially tapping the fingers on the table. The assessment of MMs was performed with the participant seated comfortably at a table with the forearms resting on the table surface or at an elevated rim, providing space to move the hands freely. The participants were instructed to execute each repetitive task at a natural speed for 10-15 seconds with each hand individually, while the other hand was resting. Thereafter, the next task was performed with no break between tasks. Thus, the whole assessment lasted up to three minutes at the maximum. All tasks were video recorded with the video camera placed in front of or above the participant orthogonally to the table surface. The video-recordings captured the participant's forearm and hands and were used to rate MMs for the resting hand according to the five levels of the W&T scale: ´0´ indicates no clear imitative movements, ´1´ suggests barely discernible repetitive movements, ´2´ slight mirror movements or stronger, but briefer, repetitive movements, ´3´ strong and sustained repetitive movements and ´4´ indicates movements equal to those observed in the active hand.[1]

Mirror movement scoring procedure

Three raters scored MMs according to the W&T scale. The raters were certified as occupational therapists (rater A and B) or physiotherapist (rater C) and participated in a training period to familiarize themselves with the W&T scale. During the rating period the raters had access to videos exemplifying the five W&T scale levels. In addition, they individually scored MMs from ten video-recordings not included in the study, prior to meeting to discuss the W&T scorings. After the training period, the raters individually scored the videos from the included participants twice with a minimum of two weeks between the ratings (rating 1 and 2). For each participant, three tasks were scored for the left and right hand separately, with scores ranging from 0 to 4 for each task. To minimize the risk of recall bias, two sets of the video-recordings were created using a random order generator so that each participant received a different order number at rating 1 and 2. The key for coupling the order

numbers for rating 1 and 2 to each participant's study identification number was stored locally in a safe digital location at Norwegian University of Science and Technology in Trondheim. None of the raters accessed this location during the rating period.

Data analyses

We investigated inter- and intrarater reliability for the sum of W&T scores for both hands for three tasks (scores ranging from 0 to 24), for the more-affected and less-affected hand separately (scores ranging from 0 to 12), and for each task for each hand separately (scores ranging from 0 to 4). In addition, separate analyses were performed for the variants of tasks 2 and 3. We used a two-way mixed model regression with the classification level as dependent variable, the test replications (rating 1 and 2) as fixed factor, and raters and participants as crossed random effects.[23] This model takes account of possible differences between raters and between rating 1 and 2. The interrater reliability Intraclass Correlation Coefficient (ICC) was calculated as the between participants (subjects) variance, divided by the total variance, see Gwet.[24] Furthermore, the intrarater reliability ICC was calculated as the sum of the between participants variance and between raters variance, divided by the total variance, see Gwet.[24]

Results

A total of 68 participants with UCP, MACS levels I-III, were recruited; 19 from Norway, 23 from Australia, and 26 from Belgium, see Table 1. Most participants were classified with MACS levels I (29%) and II (68%). Two participants were classified with MACS level III (3%). The mean AHA units were similar for the Norwegian and Belgium participants (mean AHA units: 66 and 67), while the Australian participants scored significantly lower (mean AHA unit: 52, p=0.01). The mean age for all participants was 12 years (y) and 2 months (m) (SD=3y 6m), for Norwegian participants 15 years and 11 months (SD=2y 5m), Australian participants 10 years and 5 months (SD=2y 6m), and Belgian participants 11 years (SD=2y 8m). There was one missing value at rating 1 (n= 67), while rating 2 was complete (n= 68).

The mean overall scores (total score of three tasks) and standard deviations (SD) for each combination of time points (rating 1 and 2), raters, and country are shown in Figure 1.

In the mixed effect regression model, the mean overall score at rating 2 was 1.03 points (p <0.001) lower than at rating 1. The overall ICC for interrater reliability, calculated from both ratings, was 0.90, and the corresponding intrarater reliability ICC was 0.92, see Table 2. Among the variance components used to calculate the ICCs, the variance between the participants was the largest (33.44, p≤0.001). The variance between raters was small (0.74), but statistically significant (Likelihood ratio test, p<0.001) with rater B showing a tendency to rate higher than rater C who tended to rate higher than rater A, see Figure 1. Overall, the average ratings were lower for the Belgian than the Australian and Norwegian participants, but these differences were not statistically significant (p=0.14).

The inter- and intrarater reliability coefficients for the more-affected and less-affected hand for the total scores for three tasks, as well as for each task separately, are shown in Table 2. All reliability indices for the less-affected hand were somewhat higher (ICC estimates between 0.84 and 0.92) compared to the corresponding ICCs for the more-affected hand (estimates between 0.77 and 0.88). The ICCs for task 2 and 3 performed in Belgium were lower than the ICCs for the tasks performed in Norway and Australia.

Discussion

This study demonstrates that the Woods and Teuber (W&T) scale has excellent reliability as an assessment of mirror movements (MMs) in children and adolescents with unilateral cerebral palsy (UCP). The overall interrater reliability ICC was 0.90, while the corresponding intrarater reliability ICC was 0.92. This suggests that the W&T scale can be a useful instrument in research and clinical practice by providing reliable and easily accessible information regarding the degree of MMs in children and adolescents with UCP. According to our results, the estimated reliability was higher for

MMs in the less-affected hand (ICC estimates between 0.84 and 0.92) compared to the more affected hand (ICC estimates from 0.77 to 0.88). Furthermore, our results indicate that the tasks performed in Norway and Australia, in general, produced more reliable results compared to the tasks performed in Belgium that involved sequential finger movements.

The results of the present study are consistent with those from Klingels et al.[7] who reported evidence of high inter- and intrarater reliability of the W&T scale in 20 children with UCP. Overall inter- and intrarater reliability was not stated, but ICCs of ≥ 0.82 were reported for each task for each hand separately.[7] Similar results regarding interrater reliability have been reported for a modified W&T scale in 18 children with UCP (ICCs ranging from 0.82-0.91).[10]

The W&T scale has good clinical utility and is quick and easy to administer. It does not require the use of any specific equipment, which is preferable in clinical practice and research.[1] Scoring the W&T scale can be undertaken by direct observation or from short video-recordings (≤ 3 minutes), as in the current study. The use of video-recordings enables the use of blinded raters which can be advantageous, especially for research.

It is important to acknowledge that the lack of standardization of the tasks during administration may hinder comparison of results between studies, since differences in the tasks used might affect the degree of MMs.[1,4,19] It has been suggested that highly repetitive and simple motor tasks are more appropriate to assess the presence of MMs.[8] In our study, we administered tasks that differed from those used in the original study by Woods and Teuber,[1] but similar to those tasks that were used in recent research.[3,7,19] In several recent studies, the alternating supination and pronation of the hand has been substituted with fist opening and clenching, since forceful gripping movements may elicit clearer MMs.[19] This is supported by Klingels et al[7] who reported higher occurrence of MMs for the task requiring fist opening and clenching (task 1), compared to tasks requiring sequential finger opposition and finger tapping. Results reported by Zielinski et al.[19] also indicate higher sensitivity for this task (fist opening and clenching, task 1) compared to the tasks involving sequential finger opposition and finger tapping. These sensitivity results are consistent with our results and support continued use of task 1. Nonetheless, the validity of the various tasks should be further investigated to help guide consistent administration of the W&T classification in children with UCP, including whether only using task 1 (fist opening and clenching) is sufficient for detection of MM.

Furthermore, guidelines for training and administration may ensure more reliable results when using the W&T scale. In the present study, the raters completed training prior to undertaking assessment by using video examples showing varying degrees of MMs and discussing the W&T classification levels. This may have positively affected the high reliability indices. If video-recordings are used for assessment, it is also important to ensure high quality recordings. In the present study, there seems to be more consistent scorings between raters for the participants included in Norway. This could be due to the quality of the video-recordings. All videos from Norway were recorded above the participant orthogonally to the table surface using a standardized distance from the camera to the table surface. Furthermore, downlights (from ceiling) were used to minimize possible shadows from hands when recording hand movements.

While the results from this study demonstrate that the W&T scale is a highly reliable tool for scoring MMs in children with UCP, it remains unclear whether or not the scale is capable of detecting changes in MMs following intervention or during development.[14] The W&T uses an ordinal scale with five relatively broad scoring levels that are determined from subjective observations of MMs based on their similarity with the intended movement.[1] If movement range is restricted due to muscle overactivity, stiffness or weakness, it may be more difficult to detect MMs in the more-affected hand. The somewhat higher ICCs for the less-affected hand, reported in our study and in the study by Klingels et al.,[7] provide support for this hypothesis. Laboratory based assessments use grip-force devices to measure degree of MMs, which may present more objective, less rater-dependent interval-level data. Thus, these assessments may be more sensitive in determining subtle changes

following interventions.[18,19] Accordingly, the W&T scale may be most suitable for determining the presence and degree of MMs, whereas the advantage of quantitative tools using interval-level data is that they are more objective, less rater-dependent and may more easily detect intensity levels of MMs.[19,18]

Limitations

In this study we included videos from studies previously undertaken across three different countries. The use of somewhat different tasks in Belgium compared to Norway and Australia possibly affected our main results. The separate analyses for tasks 2 and 3 show lower ICCs for the tasks performed only in Belgium. Thus, it is possible that our reliability indices would have been even higher if the tasks performed in Norway and Australia were performed by all participants. In future studies, we recommend using the tasks performed in Norway and Australia, as they have better reliability. Furthermore, our study population was not representative for the entire UCP population. Population based studies indicate that about 87% of children with UCP are classified at MACS levels I-II, while about 13% are classified at MACS level III. In our study, only two children (3%) at MACS level III were included. Although, MACS levels do not reveal whether a child has MMs or not,[7] it would have been advantageous for the generalizability of our results to have more children with MACS level III included in our study. Moreover, future research is needed to investigate the reliability of the W&T scale of children in MACS III.

Conclusion

This study demonstrates high inter- and intrareliability of the W&T scale for scoring MMs in children and adolescents with UCP. The observation-based W&T scale has excellent clinical utility and does not require specific equipment making it a feasible and useful instrument for scoring MMs in research and in clinical practice. Standardized administration guidelines for the W&T scale need to be developed to enable comparison of outcomes in future research.

Acknowledgement

The authors acknowledge the contributions from Hilde Feys, Torstein Vik, Angelina Lied and Anna Schmidt in the data acquisition.

References

1 Woods BT, Teuber HL. Mirror movements after childhood hemiparesis. *Neurology* 1978; **28**: 1152-7.

2 Staudt M, Gerloff C, Grodd W, Holthausen H, Niemann G, Krageloh-Mann I. Reorganization in congenital hemiparesis acquired at different gestational ages. *Ann. Neurol.* 2004; **56**: 854-63.

3 Kuhtz-Buschbeck JP, Sundholm LK, Eliasson AC, Forssberg H. Quantitative assessment of mirror movements in children and adolescents with hemiplegic cerebral palsy. *Dev Med Child Neurol* 2000; **42**: 728-36.

4 Koerte I, Eftimov L, Laubender RP, Esslinger O, Schroeder AS, Ertl-Wagner B, Wahllaender-Danek U, Heinen F, Danek A. Mirror movements in healthy humans across the lifespan: effects of development and ageing. *Dev Med Child Neurol* 2010; **52**: 1106-12.

5 Adler C, Berweck S, Lidzba K, Becher T, Staudt M. Mirror movements in unilateral spastic cerebral palsy: Specific negative impact on bimanual activities of daily living. *Eur J Paediatr Neurol* 2015; **19**: 504-9.

6 Friel KM, Williams PT, Serradj N, Chakrabarty S, Martin JH. Activity-Based Therapies for Repair of the Corticospinal System Injured during Development. *Front Neurol* 2014; **5**.

7 Klingels K, Jaspers E, Staudt M, Guzzetta A, Mailleux L, Ortibus E, Feys H. Do mirror movements relate to hand function and timing of the brain lesion in children with unilateral cerebral palsy? *Dev Med Child Neurol* 2016; **58**: 735-42.

8 Jaspers E, Byblow WD, Feys H, Wenderoth N. The Corticospinal Tract: A Biomarker to Categorize Upper Limb Functional Potential in Unilateral Cerebral Palsy. *Front Pediatr* 2015; **3**: 112.

9 Zielinski IM, Green D, Rudisch J, Jongsma ML, Aarts PB, Steenbergen B. The relation between mirror movements and non-use of the affected hand in children with unilateral cerebral palsy. *Dev Med Child Neurol* 2017; **59**: 152-9.

10 Riddell M, Kuo HC, Zewdie E, Kirton A. Mirror movements in children with unilateral cerebral palsy due to perinatal stroke: clinical correlates of plasticity reorganization. *Dev Med Child Neurol* 2019; **61**: 943-9.

11 Rich TL, Nemanich S, Chen CY, Sutter EN, Feyma T, Krach LE, Gillick BT. Ipsilateral Corticospinal Tract Excitability Contributes to the Severity of Mirror Movements in Unilateral Cerebral Palsy: A Case Series. *Clin EEG Neurosci* 2020; **51**: 185-90.

12 Holmstrom L, Vollmer B, Tedroff K, Islam M, Persson JK, Kits A, Forssberg H, Eliasson AC. Hand function in relation to brain lesions and corticomotor-projection pattern in children with unilateral cerebral palsy. *Dev Med Child Neurol* 2010; **52**: 145-52.

13 Islam M, Gordon AM, Skold A, Forssberg H, Eliasson AC. Grip force coordination during bimanual tasks in unilateral cerebral palsy. *Dev Med Child Neurol* 2011; **53**: 920-6.

14 Adler C, Hessenauer M, Lipp J, Kunze S, Geigenberger C, Horning A, Schaudeck M, Berweck S, Staudt M. Learning to cope with mirror movements in unilateral spastic cerebral palsy: a brief report. *Dev Neurorehabil* 2019; **22**: 141-6.

15 Kuo HC, Friel KM, Gordon AM. Neurophysiological mechanisms and functional impact of mirror movements in children with unilateral spastic cerebral palsy. *Dev Med Child Neurol* 2018; **60**: 155-61.

16 Kuhnke N, Juenger H, Walther M, Berweck S, Mall V, Staudt M. Do patients with congenital hemiparesis and ipsilateral corticospinal projections respond differently to constraint-induced movement therapy? *Dev Med Child Neurol* 2008; **50**: 898-903.

17 Simon-Martinez C, Mailleux L, Hoskens J, Ortibus E, Jaspers E, Wenderoth N, Sgandurra G, Cioni G, Molenaers G, Klingels K, Feys H. Randomized controlled trial combining constraint-induced movement therapy and action-observation training in unilateral cerebral palsy: clinical effects and influencing factors of treatment response. *Ther Adv Neurol Disord* 2020; **13**: 1756286419898065.

18 Jaspers E, Klingels K, Simon-Martinez C, Feys H, Woolley DG, Wenderoth N. GriFT: A Device for Quantifying Physiological and Pathological Mirror Movements in Children. *IEEE Trans. Biomed. Eng.* 2018; **65**: 857-65.

19 Zielinski IM, Steenbergen B, Schmidt A, Klingels K, Simon Martinez C, de Water P, Hoare B. Windmill-task as a New Quantitative and Objective Assessment for Mirror Movements in Unilateral Cerebral Palsy: A Pilot Study. *Arch. Phys. Med. Rehabil.* 2018; **99**: 1547-52.
20 Majnemer A. Benefits of using outcome measures on pediatric rehabilitation. *Phys Occup Ther Pediatr* 2010; **30**: 165-7.
21 Louwers A, Beelen A, Holmefur M, Krumlinde-Sundholm L. Development of the Assisting Hand Assessment for adolescents (Ad-AHA) and validation of the AHA from 18 months to 18 years. *Dev Med Child Neurol* 2016.
22 Eliasson AC, Krumlinde-Sundholm L, Rosblad B, Beckung E, Arner M, Ohrvall AM, Rosenbaum P. The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. *Dev Med Child Neurol* 2006; **48**: 549-54.
23 Rabe-Hesketh S, Skrondal, A. *Multilevel and Longitudinal Modeling Using Stata.* College Station, Texas: Stata Press; 2008.
24 Gwet KL. *Handbook of inter-rater reliability : the definitive guide to measuring the extent of agreement among raters.* Gaithersburg, MD: Advanced Analytics, LLC; 2012.

Tables and figures

Table 1: Demographic information for participants in terms of mean age, affected hand, gender, and functional use of the hands described with the Manual Ability Classification System (MACS) and the Assisting Hand Assessment (AHA). In addition, mean Woods and Teuber (W&T) scores for the participants are shown.

	Participating country			
	Norway	**Australia**	**Belgium**	**Total**
Participants n (%)	19 (28%)	23 (34%)	26 (38%)	68 (100%)
Mean age (SD)	16y3m (1y9m)	10y1m (2y1m)	11y (2y7m)	12y2m (3y5m)
Right more-affected hand n (%)	10 (63%)	17 (74%)	14 (54%)	41 (60 %)
Male n (%)	9 (47%)	10 (43%)	12 (46%)	31 (46 %)
MACS n (%)				
I	4 (21%)	5 (22%)	11 (42%)	20 (29%)
II	14 (74%)	18 (78%)	14 (54%)	46 (68%)
III	1 (5%)	0	1 (4%)	2 (3%)
Mean AHA unit (SD)	66.2 (16.3)	52.2 (17.0)	66.9 (19.7)	61.7 (19.0)
Mean W&T score (SD)				
Rater A				
Rating 1	8.3 (9.5)	7.5 (5.3)	4.8 (4.0)	6.7 (6.5)
Rating 2	7.8 (9.3)	6.4 (5.5)	4.7 (4.3)	6.2 (6.4)
Rater B				
Rating 1	9.1 (7.4)	9.4 (4.5)	7.8 (4.1)	8.7 (5.3)
Rating 2	8.3 (7.8)	8.4 (4.5)	7.0 (3.9)	7.9 (5.4)
Rater C				
Rating 1	8.7 (9.6)	9.8 (5.9)	6.0 (4.2)	8.1 (6.5)
Rating 2	7.5 (8.0)	7.6 (5.1)	4.4 (4.0)	6.4 (5.8)

n=number; SD=Standard Deviation; y=years; m=months.

Figure 1. Mean total scores and 95% confidence intervals (CI) of Woods and Teuber (W&T) for both hands for the three tasks for each combination of time point (Rating 1 and 2), rater (A, B, C) and country (Norway, Australia and Belgium).

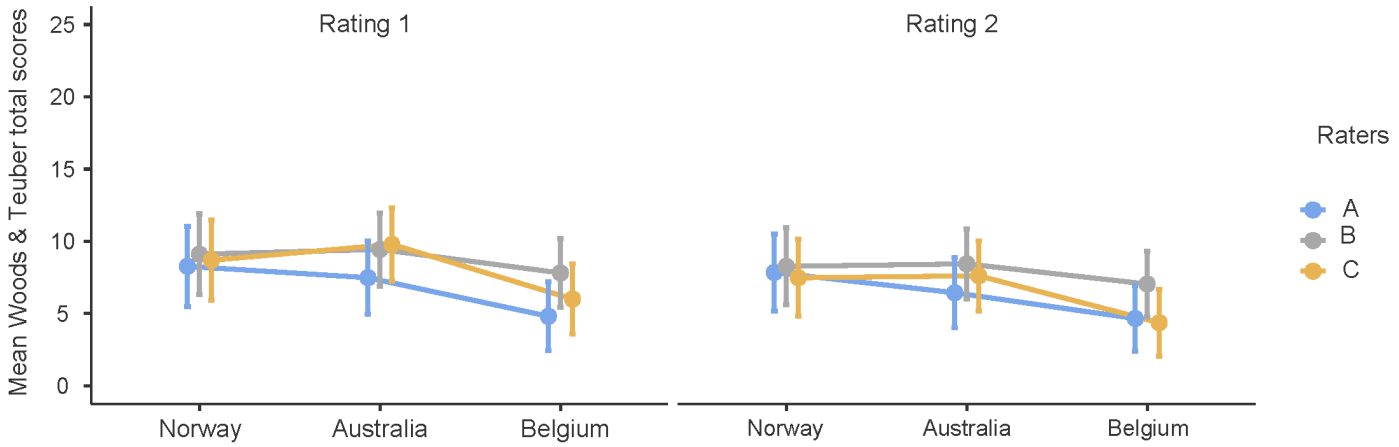

Table 2. Intraclass correlation coefficients (ICCs) for the Woods & Teuber when classifying mirror movements in the more-affected and less-affected hand. The table shows the variance components included in the equations as well as the resulting interrater- and intrarater reliability, in terms of intraclass correlation coefficients (ICCs), for the Woods and Teuber total score (overall inter- and intrarater reliability), for the more-affected and less-affected hand and for each task.

	Total score	More-affected hand				Less-affected hand			
Variance component		Total	Task 1	Task 2	Task 3	Total	Task 1	Task 2	Task 3
Between subjects	33.44	8.53	1.29	1.06	1.01	10.83	1.38	1.25	1.28
Between raters	0.74	0.19	0.03	0.01	0.03	0.18	0.02	0.02	0.02
Residual	2.93	1.20	0.23	0.23	0.27	0.95	0.18	0.20	0.23
Total	37.11	9.93	1.54	1.31	1.31	11.95	1.58	1.47	1.53
Reliability ICC									
Interrater	0.90	0.86	0.84	0.81	0.77	0.91	0.87	0.85	0.84
Intrarater	0.92	0.88	0.85	0.82	0.80	0.92	0.89	0.86	0.85

Table 3. The table displays variance components included in the equations as well as the resulting inter- and intrarater reliability, in terms of intraclass correlation coefficients (ICCs), for the various Woods and Teuber tasks that were performed.

	More-affected hand					Less-affected hand				
	All	Norway and Australia		Belgium		All	Norway and Australia		Belgium	
Variance component	Task 1	Task 2	Task 3	Task 2*	Task 3*	Task 1	Task 2	Task 3	Task 2*	Task 3*
Between subjects	1.29	1.39	1.14	0.51	0.73	1.38	1.48	1.51	0.48	0.73
Between raters	0.03	0.01	0.01	0.05	0.08	0.02	0.02	0.01	0.03	0.04
Residual	0.23	0.21	0.26	0.25	0.25	0.18	0.23	0.22	0.17	0.23
Total	1.54	1.61	1.42	0.80	1.07	1.58	1.72	1.74	0.68	1.01
Reliability ICC										
Interrater	0.84	0.86	0.81	0.63	0.69	0.87	0.86	0.87	0.71	0.73
Intrarater	0.85	0.87	0.82	0.69	0.76	0.89	0.87	0.87	0.75	0.77

Task 1= fist opening and clenching (performed in all countries)

Task 2= repetitive tapping of the index finger on the thumb (performed in Norway and Australia)

Task 3= tapping fingers simultaneously on the table (performed in Norway and Australia)

Task 2*= opposition of four fingers sequentially to the thumb (performed in Belgium)

Task 3*= sequentially tapping the fingers on the table (performed in Belgium)