Published in final edited form as:

Arch Phys Med Rehabil. 2020 November ; 101(11): 2015–2026. doi:10.1016/j.apmr.2020.05.003.

A Reaching Performance Scale for Two Wolf Motor Function Test Items

Clarisa Martinez, PT, DPT, MS¹, Helen Bacon, MS¹, Veronica Rowe, PhD, OTR/L, CBIST, FNAP², David Russak, PT, DPT, NCS, CBIS, CSCS³, Erin Fitzgerald, PT, DPT, NCS¹, Michelle Woodbury, PhD⁴, Steven L. Wolf, PT, PhD, FAPTA, FAHA, FASNR⁵, Carolee Winstein, PT, PhD, FAPTA, FAHA, FASNR¹

¹Division of Biokinesiology and Physical Therapy, University of Southern California

²Department of Occupational Therapy, Georgia State University

³Kaiser Permanente Medical Center, Woodland Hills

⁴Department of Health Sciences and Research, Medical University of South Carolina

⁵Department of Rehabilitation Medicine, Division of Physical Therapy, Emory University School of Medicine

Abstract

Objective: To adapt the Reaching Performance Scale for Stroke (RPSS) for the Wolf Motor Function Test (WMFT) “Lift Can” (*Can*) and “Hand to Box” (*Box*) items.

Design: Retrospective analysis of video-recorded WMFT assessment performed by three raters on two occasions

Setting: N/A

Participants: 29 participants with mild to moderate upper extremity impairment, <3 months post stroke

Interventions: N/A

Main Outcome Measures: Inter- and intra-rater agreement, concurrent validity of Wolf Motor Function Test-Reaching Performance Scale for Stroke (WMFT-RPSS)

Results: Mean (SD) inter-rater Gwet’s agreement coefficient (AC₂) was 0.61 (0.05) for *Can* WMFT-RPSS and 0.56 (0.03) for *Box*. Mean (SD) intra-rater AC₂ for *Can* was 0.63 (0.05) and

Corresponding Author and Reprint Contact: Clarisa Martinez, PT, DPT, MS, USC Division of Biokinesiology and Physical Therapy, 1540 Alcazar Street, CHP 155, Los Angeles, CA 90089-9006, (323) 442-2900, clarisa@pt.usc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Presentation: The visualization plots were presented as part of a Symposium, *Moving Technology to Clinical Practice* at the Combined Sections Meeting of the American Physical Therapy Association, January, 2019 in Washington DC.

Conflicts of interest: There are no conflicts of interest

Clinical Trial Registration Number Retrospective analysis of data collected under clinicaltrials.gov Identifier: [NCT00871715](https://clinicaltrials.gov/ct2/show/study/NCT00871715)

0.70 (0.04) for *Box*. WMFT-RPSS *Can* and *Box* scores correlated with log mean WMFT time (C -0.73 , B -0.48), Functional Ability Scale (C 0.87 , B 0.62), Upper Extremity Fugl-Meyer Motor Score (C 0.69 , B 0.51), and item movement rate (C 0.74 , B 0.71), (all, $p < 0.05$). Mean (SD) WMFT-RPSS score across the 29 participants was 12.7 (3.5) for *Can* (max score = 19) and 11.4 (3.0) for *Box* (max = 16).

Conclusions: WMFT-RPSS demonstrated moderate intra-rater and weak-to-moderate interrater agreement for individuals with mild-moderate impairment. For construct validity, *Can* and *Box* WMFT-RPSS were significantly correlated with four standardized measures. Average WMFT-RPSS scores revealed that some participants may have relied on compensatory movements to complete the task---a revelation not discernable from movement rate alone. The WMFT-RPSS is potentially useful as a valid and reliable tool to examine longitudinal changes in movement quality after stroke.

Keywords

Stroke; Movement Performance; Upper Extremity; Wolf Motor Function Test; Reaching Performance Scale for Stroke

The Stroke Recovery and Rehabilitation Roundtable identified a research priority to develop standardized measures distinguishing motor compensation from restitution.¹ Most upper limb ordinal-level or time-based measures like the Wolf Motor Function Test (WMFT)² lack sufficient resolution to achieve this distinction, because they do not characterize movement coordination across joints.³ Although 3D motion capture represents the gold standard for quantifying movement, kinematic measures may not adequately address both movement coordination and overall task performance.⁴ However, the Reaching Performance Scale for Stroke (RPSS),⁵ is a valid, reliable observational kinematic tool to characterize movement coordination and overall task performance during reaching. The RPSS may have further utility if its rating criteria are applied to other assessments, such as the WMFT.⁴

In this brief report, we demonstrate an approach to extract movement coordination information from WMFT performance through modification of the RPSS and application to two WMFT items. We tested the modified RPSS' reliability and validity using video and assessment data previously acquired and stored in a de-identified database for a recently completed phase III randomized clinical trial.⁶ Further, we demonstrate the potential application and usefulness of the WMFT-RPSS tool through visualization with a multi-dimensional graphical representation at the WMFT-RPSS element level that highlights inter-individual differences in movement coordination.

Methods

This is a secondary analysis from the Interdisciplinary Comprehensive Arm Rehabilitation Evaluation (ICARE, [NCT00871715](#)) Trial to test the efficacy of a structured, task-oriented training program compared to usual care.⁶ All research procedures were approved by the Institutional Review Boards (IRB) at all participating institutions. A detailed description of the RPSS modification, participant selection, and statistical analysis is found in the Expanded Methods, Appendix 1.

Wolf Motor Function Test (WMFT)

WMFT is reported as average time to perform 15 goal-directed movements. Item movement time was recorded with a stopwatch in seconds, with “120+” denoting that the task could not be performed in two minutes. The Functional Ability Scale (FAS) is a six-point ordinal rating scale scored from video for each goal-directed movement. WMFT *Can* and *Box* tasks are video recorded with one camera placed in the sagittal plane.

Reaching Performance Scale for Stroke (RPSS)

RPSS characterizes compensatory movements occurring at the trunk, shoulder, elbow, and hand during two reach-to-grasp tasks.⁵ The tool consists of six elements with ordinal rating scales (0-3). The RPSS is reported as the total of the element scores. RPSS is video recorded with one camera placed at an angle to best capture the frontal and sagittal planes.

RPSS modification for WMFT (WMFT-RPSS)

To adapt RPSS to WMFT, first, we selected two WMFT items: the “Lift Can” (*Can*) and “Hand to Box” (*Box*) items. *Can* was selected because it requires forward reach with cylindrical grasp, much like the RPSS cone items. *Box* requires forward reach without grasp and was chosen to address a possible floor effect for impaired individuals who lack distal control. RPSS ordinal grading criteria and element descriptions were modified based on the WMFT item characteristics, and to mitigate misinterpretation of the grading criteria (see Appendix 1 for details). For example, the title for the “Global Score” element from the original RPSS was changed to “Task Completion Score” to align the title with the scoring criteria, which is focused on the extent to which the task is completed. See Figure 1 for the final WMFT-RPSS rating criteria. Maximum total scores for the WMFT-RPSS *Can* and *Box* items are 19 and 16 points, respectively, and indicate that no movement compensations were observed during task performance.

Participants

Participants with mild to moderate arm hemiparesis (Upper Extremity Fugl Meyer, UEFM, 19-58) were enrolled in the ICARE trial 14-106 days after hemorrhagic or ischemic stroke at seven US hospitals, and randomized to one of three intervention groups. The intervention period was ten weeks in length, with motor assessment (including WMFT) performed by blinded, standardized evaluators at baseline, end of intervention, six months and, one year. All research procedures were approved by the Institutional Review Boards (IRB) at all participating institutions.

The Data Management and Analysis Committee (DMAC) which oversaw the ICARE trial selected de-identified demographic, assessment, and WMFT video data from 29 participants for this secondary analysis. This subset was selected because they contained complete baseline data, but lacked at least one follow up visit, thereby preserving those with complete datasets for future longitudinal analyses.

WMFT-RPSS Reliability

To test reliability, a three-person rater panel—comprised of two physical therapists and one occupational therapist with at least 2 years of clinical experience—was assembled and trained. Raters independently scored each participant's performance on the two WMFT items from video on two separate occasions (Rounds 1 and 2), separated by a three-week washout period. Raters were blinded to the timepoint at which the WMFT was administered, although all data were from baseline. We presented participant *Can* and *Box* videos separately and dispersed randomly to reduce practice effects.

Gwet's agreement coefficient (AC_2)⁷ was used to estimate inter- and intra-rater agreement for the total WMFT-RPSS score and for individual WMFT-RPSS elements. AC_2 is appropriate to test multi-rater agreement for ordinal level data and is weighted to penalize larger differences in scores more than smaller differences. Interpretation of AC_2 is similar to kappa, such that rater agreement can be characterized as “None” (0-0.2), “Minimal” (0.21-.39), “Weak” (0.4-0.59), “Moderate” (0.6-0.79), “Strong” (0.8-0.9), and “Almost Perfect” (0.9+) agreement.⁸

WMFT-RPSS Concurrent Validity

To test concurrent validity, we used Spearman (r_s) rank correlation coefficients to compare WMFT-RPSS to log mean WMFT time, mean FAS, and Upper Extremity Fugl-Meyer (UEFM) Motor Score. To conduct these correlation analyses, we tabulated all rater data across the two scoring rounds and derived a composite WMFT-RPSS score for each participant based on the element scores that had the highest frequency counts.

We planned to determine the correlation between item-level movement time (s) and total WMFT-RPSS score. However, the “120+” time score is problematic because participants who could not perform a task do not have a time score.⁹ In lieu of WMFT item movement time, we used *item rate*, calculated as $60(s)/TaskPerformanceTime(s)$. For individuals with a time score more than 120 seconds, the movement rate was recorded as 0 reps/60s to indicate that they performed zero task repetitions within 60 seconds.⁹ Descriptive statistics, reliability and validity were calculated using SAS Studio 3.8.

WMFT-RPSS Application

To visualize inter-individual differences at the element score level as they relate to task performance, we plotted polygon graphs of WMFT-RPSS element scores superimposed at the respective intersections of total WMFT-RPSS (Y-axis) and item movement rate (X-axis). We provide descriptive statistics of participants' movement characteristics for both tasks.

Results

Sample characteristics

All participants could perform the *Box* item, but two could not perform the *Can* item. This was consistent with inclusion criteria for the parent project, a cohort with mild to moderate stroke impairment. See Appendix 2 for demographic characteristics of the 29 participants.

Inter- and intra-rater agreement

Detailed results of the inter- and intra-rater agreement analyses are found in Appendix Mean (SD) inter-rater AC₂ across both Rounds was 0.61 (0.05) for *Can* WMFT-RPSS, and 0.56 (0.03) for *Box*. Mean (SD) AC₂ was highest for the *Can* and *Box* “Task Completion” element [0.81 (0.05) and 0.68(0.02), respectively], and lowest for *Can* “Shoulder Movement” [0.53 (0.03)] and *Box* “Movement Continuity” [0.45 (0.03)]. Mean (SD) intra-rater AC₂ for the *Can* WMFT-RPSS score was 0.63 (0.05), and 0.70 (0.04) for the *Box* WMFT-RPSS score. Mean intra-rater AC₂ was highest for *Can* “Task Completion” [0.83 (0.09)] and *Box* “Trunk Displacement” [0.78 (0.08)] and lowest for *Can* and *Box* “Movement Continuity” [0.56 (0.17) and 0.63 (0.02), respectively].

Concurrent validity

Can WMFT-RPSS demonstrated significant correlations with UEFM ($r_s=0.69$, $p<0.001$), mean WMFT FAS ($r_s=0.86$, $p<0.0001$), and log mean WMFT time ($r_s=-0.73$, $p<0.0001$). *Box* WMFT-RPSS demonstrated significant correlations with UEFM ($r_s=0.51$, $p<0.0048$), FAS ($r_s=0.62$, $p<0.001$), and log mean WMFT time ($r_s=-0.48$, $p<0.01$). There was a significant correlation between item movement rate and WMFT-RPSS for *Can* ($r_s=0.74$, $p<0.0001$) and *Box* ($r_s=0.70$, $p<0.0001$).

WMFT-RPSS Application

The polygon graphs in Figure 2A and B detail inter-individual element scores as they relate to movement time and WMFT-RPSS. The figure highlights case examples in which participants have the same movement rate but different WMFT-RPSS scores (2A), and participants with the same WMFT-RPSS score but different element score patterns and movement rates (2B). Average (SD) WMFT-RPSS scores for *Can* and *Box* were 12.7 (3.5) and 11.4 (3.0), respectively. Two participants for each item were rated with the maximum score, indicating no compensatory movement strategies were observed. No participants received the maximum total score for both items. Participants most frequently received the maximum score on the *Can* “Trunk Stability” element (41.4%), and *Box* “Task Continuity” element (41.4%).

Discussion

We demonstrated an approach to characterize WMFT task performance for individuals with mild to moderate stroke by adapting the RPSS. Intra-rater agreement for total WMFT-RPSS score was moderate, and weak-to-moderate for inter-rater agreement. Possible contributing factors to low rater agreement may be the revised wording of the adapted tool, or that raters viewed the movement in the sagittal plane compared to the angled camera position used in the original RPSS. The WMFT-RPSS was significantly correlated with four other validated standardized measures, although r_s were generally higher for *Can* than *Box*. From the multidimensional polygon plots, we observed that patterns of element scores were variable, even among participants with similar WMFT-RPSS scores or movement times. The WMFT-RPSS may be useful to detect differences in task performance among individuals with the same movement time. Although most participants demonstrated some compensation, we also detected a small subset whose movement pattern lacked compensation as indicated by a

maximum WMFT-RPSS score at baseline. The clinical value of identifying individuals with more restored movement performance is that treatment approaches for this subset may differ from those who enlist compensatory movement strategies.

Study Limitations

Limitations include the lower AC_2 of some elements and the stopwatch method used to determine movement time.

Conclusions

We presented a method to rate movement performance quality in persons with mild-moderately impaired hemiparesis for two WMFT items. Future work is required to improve WMFT-RPSS inter-rater reliability. An improved WMFT-RPSS tool would have the potential to track changes over time and/or to inform more efficient and personalized rehabilitation for stroke survivors.

Acknowledgements

The authors would like to thank Beth Fisher, Nicolas Schweighofer, Christianne Lane, Caron Park, Li Ding, Monica Nelsen, and Bruce Dobkin for helpful comments on an earlier version of the manuscript. We also would like to thank Bokkyu Kim for his contribution to the multidimensional polygon plot visualization.

Financial Support:

Research reported in this publication was supported by the Eunice Kennedy, Shriver National Institute of Child Health & Human Development (NICHD), the, National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health, and the National Center for Advancing Translational Science (NCATS) under Award Numbers U01NS056256, R01HD065438, T32HD064578, UL1TR001855, and UL1TR000130. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix 1:: Expanded Methods

Wolf Motor Function Test Data Acquisition

This is a secondary analysis of WMFT data collected as part of the Interdisciplinary Comprehensive Arm Rehabilitation Evaluation study (ICARE).^{1,2} The ICARE study was prospectively registered on clinicaltrials.gov (identifier: [NCT0087175](https://clinicaltrials.gov/ct2/show/study/NCT0087175)). All research procedures were approved by the Institutional Review Boards (IRB) at all participating institutions: University of Southern California Health Sciences Campus (IRB#HS-07-00148), Emory University (IRB#IRB00001180), MedStar National Rehabilitation Hospital/MedStar Health Research Institute (IRB#2007-161), Rancho Los Amigos National Rehabilitation Center (IRB #037), Casa Colina Centers For Rehabilitation (IRB #HS-07-00148), Long Beach Memorial Care Health System (IRB #428-07), Huntington Memorial Hospital Institutional Review Board (IRB #HMH 2007049), and Cedars-Sinai (IRB #Pro00012032).

Participants were recruited from the seven participating institutions, typically during inpatient rehabilitation. In brief, individuals were eligible for enrollment if they had subacute stroke (14-106 days after stroke), were age 21 or older, with ischemic or hemorrhagic stroke

and upper extremity hemiparesis, voluntary finger extension, and if they had not received more than six therapy sessions in an outpatient setting. Participants were excluded if other traumatic or nonvascular brain injuries were present, or if they had subarachnoid or primary ventricular hemorrhage (for a full list of inclusion and exclusion criteria, please see Winstein et. al, 2016^{1,2}). Participants were randomized to one of three groups: an experimental task oriented motor training program, a usual and customary care program, and usual and customary care program that was dose- matched to the task oriented training program.

The WMFT was the primary outcome, measured at four timepoints during the year-long intervention and follow up period (baseline, end of intervention, six months, one year). Blinded ICARE assessors underwent standardization and routine auditing to ensure that WMFT was administered consistently at all five study sites.

Detailed methods of WMFT task setup, video recording, and participant instructions for the ICARE study are described by Duff and colleagues.³ The WMFT was recorded by a camera placed in the sagittal plane, using a fixed camera distance and detailed framing criteria. Assessors provided participants with verbal instructions and demonstrated the task; participants were instructed to not practice the task. Participants were allowed up to two attempts for each task. The Lift Can and Hand to Box tasks were among the 15 timed items of the WMFT. These were selected from among the 15 given their similarity to the RPSS cone task (see Brief Report for rationale).

Reaching Performance Scale for Stroke (RPSS)

The RPSS was created by Levin and colleagues⁴ as an observational kinematics tool to describe compensatory movements of the paretic arm and hand. The tool is used to rate two reach-to-grasp tasks. Participants are seated at a table, with a cylindrical cone placed at midline of the participant's body. For one task—a “near” reach—the cone is 1 cm from the table's edge and for the second task (“far”), the cone is 30 cm from the edge. Performance was video recorded with one camera placed on the paretic side, and at an angle to capture both frontal and sagittal planes. The movements were scored from video and raters only scored the “transport” phase of the reach. The RPSS has 6 ordinal items to score reaching performance, described in detail in their paper.⁴ There are two sets of rating criteria for the “near” and “far” tasks.

The RPSS demonstrated concurrent validity with several measures of paretic limb function, such as grip force, Chedoke-McMaster Stroke Assessment and The Upper Extremity Performance Test for the Elderly (TEMPA) and had mean interrater agreement of 67% (SD=13.5%). The authors recommend that future work could include improving the rating scale descriptions to increase rater agreement.

RPSS modification for WMFT (WMFT-RPSS)

A panel of clinician-researchers from the ICARE study team adapted the RPSS to address areas where there was confusion regarding the rating criteria interpretation and to make the scale applicable to the two WMFT tasks. After completing the tool modifications, we selected a new three-person rating panel whose members did not participate in the tool adaptation process. The panel of two physical therapists and one occupational therapist were

experienced clinicians with at least 2 years of clinical experience. They participated in seven one-hour long training sessions during which time they independently rated movement performance for the Can and Box tasks for a set of practice videos. Questions from the new panel members regarding item ambiguity were noted, and further changes to the rating criteria and scoring instructions were made as needed.

The differences between the original RPSS and the WMFT-RPSS are summarized in Appendix 1, Table 1 below. The final RPSS tool is presented in Figure 1 of the Brief Report.

Appendix 1, Table 1:

WMFT-RPSS modifications from original RPSS scoring criteria

Reaching Performance Scale for Stroke—RPSS (Levin and colleagues, 2004) ⁴	Wolf Motor Function Scale—Reaching Performance Scale for Stroke—WFMT-RPSS	Summary of modifications made
Trunk Displacement	Trunk displacement	
3. (Near) “No or almost not forward trunk displacement” (Far) “Appropriate forward trunk displacement related to the amount of elbow extension”	3. No or almost no forward trunk displacement relative to the other side	Added reference to the uninvolved side Removed reference to elbow extension to minimize overlap with other rating elements
2. (Near) “Small displacement of the trunk (flexion, rotation, or flexion accompanied by rotation)” (Far) “Excessive trunk displacement related to a limitation of the active movement of the elbow or shoulder”	2. Less than half the movement is made by the trunk: (flexion, rotation, or flexion accompanied by rotation)	Quantified the amount of displacement by adding “less than half” Removed reference to relationship between trunk and shoulder movement to minimize overlap with other rating elements
1. (Near) “More than half the movement is made by the trunk” (Far) “Excessive trunk displacement: about half of the displacement of the hand toward the target is accomplished by the trunk but the hand arrives at the target”	1. Initiate the movement with the trunk, or trunk completes majority of the reaching movement	Added timing of movement (initiation) Removed reference to displacement of the hand and degree of task accomplishment to minimize overlap with other rating elements
0. (Near) “Task is accomplished only by forward trunk displacement” (Far) “Excessive trunk displacement: more than three fourths of the displacement of the hand to the target is accomplished by the trunk, and the hand does not arrive at the target”	0. Movement is accomplished only by forward trunk displacement, with minimal contribution from shoulder and elbow (eg: head position moves over the table to complete the reach)	Removed reference of degree of task accomplishment to minimize overlap with other rating elements Added criteria to determine extent of trunk displacement (position of head relative to table)
Movement Smoothness	Movement Continuity	Changed title because “smoothness” typically refers to the kinematic measure of jerk, which is a derivative of position and not observable. Generally, changed all definitions for this category to refer to hand’s movement toward target rather than movement of arm and trunk to narrow scope of the rating criteria
3. “The combination of movement of the arm and trunk is fluid and smooth”	3. The hand movement is continuous	Changed “fluid and smooth” to “continuous”
2. “More than one movement of the arm is made to perform the task, or the movement is segmented (not smooth)”	2. One point at which a correction in hand movement is made	Changed rating to “one movement” instead of “more than one movement” to help distinguish a rating of “2” from a rating of “1”

Reaching Performance Scale for Stroke—RPSS (Levin and colleagues, 2004) ⁴	Wolf Motor Function Scale—Reaching Performance Scale for Stroke—WFMT-RPSS	Summary of modifications made
1. “Several small of movements of the arm and trunk are made in a sequential manner”	1. More than one correction of the hand movement is made to perform the task	Changed rating to “more than one correction of the hand...” to distinguish a rating of “2” from a rating of “1”
0. “Complete segmentation of arm and trunk movement”	0. Multiple pauses and changes in hand trajectory that do not adequately progress hand to the target	Added that pauses and changes do not adequately progress hand to target to distinguish a rating of “1” from “0”
Shoulder movement	Shoulder movement	
3. (Near) “Adequate shoulder flexion and horizontal adduction with scapular elevation to perform the task” (Far) “Adequate shoulder flexion and horizontal adduction with scapular protraction and elevation to perform the task”	3. Normal shoulder flexion and horizontal adduction with no scapular elevation, minimal protraction at the end of the reach	Changed “adequate” to “normal”, modified the description of the expected shoulder movement so that it is distinct from a rating of “2”
2. (Near) “Shoulder flexion and horizontal adduction occurs with excessive scapular elevation” (Far) “Shoulder flexion and horizontal adduction occurs with excessive scapular protraction or elevation”	2. Movement is primarily in the sagittal plane but insufficient shoulder flexion and/or excessive shoulder abduction at the end of the reach.	Detailed that movement is expected to be in the sagittal plane but is insufficient, to distinguish this rating from a “3”
1. (Near) “Shoulder flexion occurs only in combination with excessive scapular elevation. Shoulder horizontal adduction is decreased” (Far) “Shoulder flexion is combined with scapular elevation. Shoulder horizontal adduction is decreased”	1. Movement is initiated with scapular elevation, or arm elevation is a combination of shoulder abduction, internal rotation, scapular protraction with little flexion	Added criteria related to the timing of scapular elevation
0. “No or almost no shoulder flexion or horizontal adduction is possible (all the movement is made by the scapula)”	0. No shoulder movement occurs	Changed criteria to “no movement occurs” to distinguish it from a rating of “1”
Elbow Movements	Elbow Movements	
3. (Near) “Extending the hand to the target is principally attributed to elbow extension” (Far) “Elbow extension is almost full”	3. Normal elbow extension (extent of elbow extension at the end of reach to grasp is consistent with less-affected side)	Added comparison to less-affected side
2. “More than half of the reaching movement is attributed to elbow extension”	2. Active elbow extension occurs, and is a major contributor to the movement, but does not reach full extension at the end of reach to grasp relative to less-affected side.	Changed “more than half of the reaching movement...” to elbow extension is a “major contributor” compared to the less-affected side
1. “Less than half of the reaching movement is attributed to elbow extension”	1. Active elbow extension is a minor contributor to the reach to grasp movement	Changed “less than half...” to elbow extension is “minor contributor”
0. “No elbow extension occurs”	0. No or almost no elbow extension occurs	
Prehension	Prehension	Prehension will be scored for the Lift Can task only, not the Hand to Box task
3. “Adequate hand opening and closure to perform the task”	3. Normal hand opening and closure to perform the task relative to the other side	Changed from “adequate” to “normal relative to the less-affected side” to help distinguish a score of “3” from “2”
2. “Opening or relaxing the hand is difficult”	2. Adequate hand opening and closure to perform the task (correct grasp but not normal), excessive/reduced hand opening	Changed to “adequate” to distinguish a score of “2” from “3”

Reaching Performance Scale for Stroke—RPSS (Levin and colleagues, 2004) ⁴	Wolf Motor Function Scale—Reaching Performance Scale for Stroke—WMFT-RPSS	Summary of modifications made
1. “Use of compensatory grasping strategies (eg, winding fingers around a cone, downward grasping)”	1. Use of compensatory grasping strategies (eg, winding fingers around object, downward grasping)	Removed reference to a cone since the task involves a can
0. “Prehension is not possible”	0. Prehension is not possible	
Global Score	Task Completion	<p>Changed title so that the focus is on the extent to which task is completed, consistent with the rating criteria</p> <p>Generally task completion is used to rate coordination of trunk, shoulder, and elbow movement in the context of the whole task performance</p> <p>Made consistent reference to the movement relative to the less-affected side</p>
3. “The task can be done easily, with or without mild tremor or dysmetria, following a smooth and direct trajectory”	4. The task can be done easily on the first attempt; the timing and sequence of trunk/shoulder/elbow/hand movements are consistent with the less-affected side	<p>Added that task must be completed on the first attempt</p> <p>Added that timing and sequence are consistent with the less-affected side</p>
2. “The task is done in the presence of tremor; dysmetria; small, jerky movements; arc-shaped trajectory; or segmentation. Prehension is possible but may be modified or difficult”	3. The task is completed on the first attempt, but performance is not consistent with the less-affected side (arm bumps table during reach; abnormal timing or sequence of movement; tremor; dysmetria; arc-shaped trajectory)	<p>Added that the task is completed on the first attempt but not consistent with the less-affected side</p> <p>Gave WMFT-specific example (arm bumping table) that would affect scoring</p>
1. “The task is done partially (more than 50%) or with modification (such as stabilization of the cone, sliding the cone on the table, modification of table height, shorter distance to the cone). Prehension may be absent”	2. The task is completed on the first attempt, but required modification (eg arm rests on table during grasp, movement is much slower, can moves during grasp)	<p>Removed reference to prehension to distinguish it from the Prehension rating criteria</p> <p>Added examples that are more specific to the WMFT tasks rather than the RPS cone task</p>
0. “Less than half the task is accomplished despite modifications”	1. The task required several attempts for completion (eg. Several attempts to raise arm to box or can, several attempts at grasp, or adjusting position of object to grasp).	<p>Added this rating level so successful task completion after multiple attempts could also be scored</p> <p>Added possible task modifications that the raters may observe</p>
	0. Task is not completed or partially completed.	

Participant Selection

We requested access to de-identified ICARE WMFT video, demographic, and performance scores from the Data Management and Analysis Committee (DMAC), which oversees the WMFT video database for the ICARE clinical trial. The DMAC provided de-identified data equally allocated across groups if: 1) there was a written consent to be videotaped, 2) time scores, FAS scores and video were available for both can and hand to box items, and 3) upper extremity Fugl Meyer (UEFM) scores were available. To preserve complete datasets for future analysis, the DMAC further recommended that we perform our rater comparisons on a subset of data with incomplete datasets. Appendix 1, Figure 1 details the selection process from the ICARE database. Participant data from 29 individuals was used for this study.

Appendix 1, Figure 1:
WMFT-RPSS selection process.

Reliability of WMFT-RPSS

Video recordings of the Lift Can and Hand to Box tasks for both affected and less-affected sides were from the WMFT recordings. The videos were edited so that affected and less affected sides were presented together (to allow for rater comparison). However, the Can and Box tasks for each participant were separated into two video files so that the performance on one task did not influence scoring of the other.

To test inter and intra-rater reliability, each rater independently reviewed and rated the Can and Box items of 29 participants on two separate occasions, with a three-week washout period in between ratings. For the Can task, raters were instructed to score only the “transport” portion of the Lift Can task, but not the return movement. Although all participant videos were from the Baseline assessment, raters were kept blinded to the time-point at which the WMFT was administered. To limit any learning effect, the videos were presented to the raters so that the Can and Box task for all participants were randomly

ordered. Raters could pause and view the video frame by frame as they scored the movement. Raters submitted their video ratings using an online survey created through Qualtrics (<https://www.qualtrics.com/>).

Statistical analysis

Descriptive statistics were used to characterize the 29 participants. Inter- and intra-rater reliability were tested using Gwet's agreement coefficient (AC). This statistic is appropriate for ordinal measures and multiple raters, and can account for larger differences in ratings more than small.⁵ We compared ratings for each of the six elements and the total WMF-RPSS score for the Can and Box tasks to determine inter- and intra-rater reliability.

Concurrent validity was assessed using Spearman's rank correlation coefficient (r_s) to compare total WMFT-RPSS score to the UEFM⁶, the log-transformed mean WMFT time, and mean FAS score. For these comparisons, we tabulated rating scores across the three raters for both rating attempts and used the element scores with the highest frequencies. In cases where there were two-way ties, we used the lower of the two scores. There were two cases of three-way ties for the Box task Continuity element; in these cases, we used the middle of the three scores. Descriptive statistics, reliability and validity were calculated using SAS Studio 3.8.

We explored the association between WMFT-RPSS score elements and movement time using regression analysis. In order to address the issue of missing movement time under circumstances in which a participant was unable to perform a task with the paretic side, movement time was transformed to a movement rate (reps/60s), calculated as $60(s)/\text{TaskPerformanceTime}(s)$ ⁷. Consistent with Hodics and colleagues, we assigned a rate of 0 reps/60s for participants who could not perform the task within 120 seconds.

References

1. Winstein CJ, Wolf SL, Dromerick AW, et al. Interdisciplinary Comprehensive Arm Rehabilitation Evaluation (ICARE): a randomized controlled trial protocol. *BMC Neurol*. 2013;13:5. doi:10.1186/1471-2377-13-5 [PubMed: 23311856]
2. Winstein CJ, Wolf SL, Dromerick AW, et al. Effect of a task-oriented rehabilitation program on upper extremity recovery following motor stroke: the ICARE randomized clinical trial. *JAMA*. 2016;315(6):571–581. doi:10.1001/jama.2016.0276 [PubMed: 26864411]
3. Duff S V, He J, Nelsen MA, et al. Interrater reliability of the Wolf Motor Function Test-Functional Ability Scale: why it matters. *Neurorehabil Neural Repair*. 2015;29(5):436–443. doi:10.1177/1545968314553030 [PubMed: 25323459]
4. Levin MF, Desrosiers J, Beauchemin D, Bergeron N, Rochette A. Development and validation of a scale for rating motor compensations used for reaching in patients with hemiparesis: the reaching performance scale. *Phys Ther*. 2004;84(1):8–22. [PubMed: 14992673]
5. SAS Institute Inc. Compute estimates and tests of agreement among multiple raters. <http://support.sas.com/kb/25/006.html>. Published 2019 Accessed November 22, 2019.
6. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7(1):13–31. <http://www.ncbi.nlm.nih.gov/pubmed/1135616>. Accessed February 20, 2018. [PubMed: 1135616]
7. Hodics TM, Nakatsuka K, Upreti B, Alex A, Smith PS, Pezzullo JC. Wolf Motor Function Test for characterizing moderate to severe hemiparesis in stroke patients. *Arch Phys Med Rehabil*. 2012;93(11):1963–1967. doi:10.1016/j.apmr.2012.05.002 [PubMed: 22579647]

Appendix 2:

Demographic characteristics of study participants

Participant Characteristics		n=29			
Age (y)	Mean (SD)	56.7 (13.2)			
Duration from stroke (days)	Mean (SD)	44.8 (25.0)			
Sex	Freq (%)	Women	12	(41.4)	
		Men	17	(58.6)	
Stroke Type	Freq (%)	Ischemic	24	(82.8)	
		Hemorrhagic	4	(13.8)	
		Other	1	(3.4)	
Stroke Location	Freq (%)	R Hemisphere	14	(48.3)	
		L Hemisphere	14	(48.3)	
		Brainstem	1	(3.4)	
Pre-Stroke Hand Dominance	Freq (%)	Right	28	(96.6)	
		Left	1	(3.4)	
Stroke Severity	Freq (%)	Less severe (UEFM ≥ 36)	24	(82.8)	
		More severe (UEFM<36)	5	(17.2)	
Clinical Characteristics					
Upper Extremity Fugl Meyer (UEFM)	Mean (SD)	44.1 (9.7)			
Mean Wolf Motor Function Test time (WMFT, s)	Median (range)			(2.06 - 50.94)	
		More affected side	4.04	(1.36 - 18.34)	
		Less affected side	2.21		
Task Movement rate (reps/60 s)	Mean (SD)	More affected side	22.6	(11.1)	
		Less affected side	44.3	(11.9)	
		Lift Can	55.5	(24.8)	
		Hand to Box	84.9	(28.2)	
WMFT Functional Abilities Scale (WMFT-FAS, N=28)	Mean (SD)	3.5 (0.7)			
Wolf Motor Function Test-Reaching Performance Scale (WMFT-RPSS)	Mean (SD)	Lift Can (0-19)	12.7	(3.5)	
		Hand to Box (0-16)	11.4	(3.0)	

Appendix 3:

Inter-rater and Intra-rater agreement of the WMFT-RPSS

Inter-rater Reliability	Gwet's Agreement Coefficient (AC)									
Can Task WMFT-RPSS Element	Round 1					Round 2				
	Gwet's AC*	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit	Gwet's AC*	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit
Trunk displacement	0.64	0.13	<.0001	0.38	0.90	0.55	0.09	<.0001	0.37	0.73
Movement continuity	0.60	0.09	<.0001	0.43	0.77	0.56	0.10	<.0001	0.37	0.75
Shoulder movement	0.55	0.15	0.0002	0.27	0.84	0.51	0.23	0.0289	0.05	0.97
Elbow movement	0.71	0.06	<.0001	0.59	0.84	0.59	0.06	<.0001	0.48	0.71

Inter-rater Reliability	Gwet's Agreement Coefficient (AC)									
Can Task WMFT-RPSS Element	Round 1					Round 2				
	Gwet's AC*	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit	Gwet's AC*	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit
Prehension	0.77	0.06	<.0001	0.66	0.88	0.73	0.10	<.0001	0.54	0.93
Task completion	0.85	0.07	<.0001	0.70	0.99	0.78	0.07	<.0001	0.64	0.92
Total Can WMFT-RPSS score	0.64	0.05	<.0001	0.55	0.74	0.58	0.07	<.0001	0.45	0.71
Box Task WMFT-RPSS Element	Round 1					Round 2				
	Gwet's AC*	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit	Gwet's AC*	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit
Trunk displacement	0.69	0.10	<.0001	0.48	0.89	0.61	0.15	<.0001	0.32	0.90
Movement continuity	0.43	0.15	0.0054	0.13	0.73	0.48	0.23	0.0345	0.03	0.92
Shoulder movement	0.56	0.13	<.0001	0.31	0.81	0.55	0.09	<.0001	0.37	0.73
Elbow movement	0.69	0.08	<.0001	0.53	0.85	0.59	0.12	<.0001	0.35	0.82
Task completion	0.67	0.06	<.0001	0.55	0.78	0.70	0.13	<.0001	0.44	0.96
Total Box WMFT-RPSS score	0.59	0.05	<.0001	0.49	0.68	0.54	0.14	<.0001	0.27	0.81

Intra-Rater Reliability	Gwet's Agreement Coefficient (AC)									
Can Task WMFT-RPSS Element	Rater 1					Rater 2				
	Gwet's AC2	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit	Gwet's AC2	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit
Trunk displacement	0.51	0.10	<.0001	0.32	0.71	0.75	0.09	<.0001	0.58	0.93
Movement continuity	0.49	0.12	<.0001	0.26	0.73	0.43	0.12	0.0005	0.19	0.67
Shoulder movement	0.64	0.10	<.0001	0.44	0.83	0.66	0.10	<.0001	0.46	0.86
Elbow movement	0.69	0.07	<.0001	0.56	0.83	0.61	0.10	<.0001	0.42	0.81
Prehension	0.87	0.06	<.0001	0.76	0.98	0.78	0.07	<.0001	0.64	0.91
Task completion	0.89	0.04	<.0001	0.82	0.97	0.73	0.06	<.0001	0.60	0.85
Total Can WMFT-RPSS score	0.68	0.05	<.0001	0.58	0.79	0.60	0.06	<.0001	0.48	0.72
Box Task WMFT-	Rater 1					Rater 2				

Intra-Rater Reliability	Gwet's Agreement Coefficient (AC)									
Can Task WMFT-RPSS Element	Rater 1					Rater 2				
	Gwet's AC2	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit	Gwet's AC2	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit
RPSS Element	Gwet's AC2	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit	Gwet's AC2	Standard Error	p-value	Lower Confidence Limit	Upper Confidence Limit
	Trunk displacement	0.72	0.09	<.0001	0.54	0.89	0.88	0.06	<.0001	0.76
Movement continuity	0.76	0.08	<.0001	0.41	0.82	0.62	0.12	<.0001	0.39	0.85
Shoulder movement	0.62	0.10	<.0001	0.54	0.88	0.74	0.09	<.0001	0.57	0.92
Elbow movement	0.71	0.09	<.0001	0.47	0.85	0.69	0.11	<.0001	0.47	0.92
Task completion	0.66	0.10	<.0001	0.60	0.91	0.71	0.09	<.0001	0.53	0.89
Total Box WMFT-RPSS score	0.66	0.06	<.0001	0.53	0.78	0.75	0.05	<.0001	0.64	0.85

Abbreviations:

AC₂	Gwet's agreement coefficient for ordinal data
Box	Wolf Motor Function Test Hand to Box Item
Can	Wolf Motor Function Test Lift Can Item
DMAC	Data Management and Analysis Committee
FAS	Wolf Motor Function Test Functional Ability Scale
ICARE	Interdisciplinary Comprehensive Arm Rehabilitation Evaluation
NICHD	Eunice Kennedy Shriver National Institute of Child Health and Human Development
NINDS	National Institute of Neurological Disorders and Stroke
RPSS	Reaching Performance Scale for Stroke
UEFM	Upper Extremity Fugl Meyer
WMFT	Wolf Motor Function Test
WMFT-RPSS	Wolf Motor Function Test-Reaching Performance Scale for Stroke

References

1. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized Measurement of Sensorimotor Recovery in Stroke Trials: Consensus-Based Core Recommendations from the Stroke Recovery and

- Rehabilitation Roundtable. *Neurorehabil Neural Repair*. 2017;31(9):784–792. doi:10.1177/1545968317732662 [PubMed: 28934918]
2. Morris DM, Uswatte G, Crago JE, Cook EW, Taub E. The reliability of the Wolf Motor Function Test for assessing upper extremity function after stroke. *Arch Phys Med Rehabil*. 2001;82(6):750–755. doi:10.1053/apmr.2001.23183 [PubMed: 11387578]
 3. Demers M, Levin MF. Do Activity Level Outcome Measures Commonly Used in Neurological Practice Assess Upper-Limb Movement Quality? *Neurorehabil Neural Repair*. 2017;31(7):623–637. doi:10.1177/1545968317714576 [PubMed: 28675943]
 4. Tomita Y, Rodrigues MRM, Levin MF. Upper Limb Coordination in Individuals With Stroke: Poorly Defined and Poorly Quantified. *Neurorehabil Neural Repair*. 2017;31(10– 11):885–897. doi:10.1177/1545968317739998 [PubMed: 29130396]
 5. Levin MF, Desrosiers J, Beauchemin D, Bergeron N, Rochette A. Development and validation of a scale for rating motor compensations used for reaching in patients with hemiparesis: the reaching performance scale. *Phys Ther*. 2004;84(1):8–22. [PubMed: 14992673]
 6. Winstein CJ, Wolf SL, Dromerick AW, et al. Effect of a task-oriented rehabilitation program on upper extremity recovery following motor stroke: the ICARE randomized clinical trial. *JAMA*. 2016;315(6):571–581. doi:10.1001/jama.2016.0276 [PubMed: 26864411]
 7. SAS Institute Inc. Compute estimates and tests of agreement among multiple raters. <http://support.sas.com/kb/25/006.html>. Published 2019 Accessed November 22, 2019.
 8. McHugh ML. Interrater reliability: the kappa statistic. *Biochem medica*. 2012;22(3):276–282. <http://www.ncbi.nlm.nih.gov/pubmed/23092060>.
 9. Hodics TM, Nakatsuka K, Upreti B, Alex A, Smith PS, Pezzullo JC. Wolf Motor Function Test for characterizing moderate to severe hemiparesis in stroke patients. *Arch Phys Med Rehabil*. 2012;93(11):1963–1967. doi:10.1016/j.apmr.2012.05.002 [PubMed: 22579647]

- A modified RPSS scale may detect restitution from compensation in WMFT tasks
- WMFT-RPSS had moderate intra-rater and weak-moderate inter-rater agreement
- WMFT-RPSS was significantly correlated with other functional measures
- Participants demonstrated compensation not evident from task time alone

Score	Trunk Displacement	Movement Continuity	Shoulder Movement	Elbow Movement	Prehension (Can task only)	Task Completion
0	Movement is accomplished only by forward trunk displacement with minimal contribution from shoulder and elbow	Multiple pauses and changes in hand trajectory that do not adequately progress hand to the target	No shoulder movement occurs	No or almost no elbow extension occurs	Prehension is not possible	Task is not completed or partially completed
1	Initiate the movement with the trunk, or trunk completes the majority of the reaching movement	More than one correction of the hand movement is made to perform the task	Movement is initiated with scapular elevation or arm elevation is a combination of shoulder abduction, internal rotation, scapular protraction with little flexion	Active elbow extension is a minor contributor to the reach to grasp movement	Use of compensatory grasping strategies (eg, winding fingers around object, downward grasping)	The task required several attempts for completion (eg. Several attempts to raise arm to box or can, several attempts at grasp, or adjusting position of object to grasp)
2	Less than half the movement is made by the trunk (flexion, rotation or flexion accompanied by rotation)	One point at which a correction in hand movement is made	Movement is primarily in the sagittal plane but insufficient shoulder flexion and/or excessive shoulder abduction at the end of the reach	Active elbow extension occurs, and is a major contributor to the movement, but does not reach full extension at the end of reach to grasp relative to the less affected side	Adequate hand opening and closure to perform the task (correct grasp but not normal), excessive/reduced hand opening	The task is completed on the first attempt, but required modification (eg arm rests on table during grasp, movement is much slower, can moves during grasp)
3	No or almost no forward trunk displacement relative to the other side	The hand movement is continuous	Normal shoulder flexion and horizontal adduction with no scapular elevation, minimal protraction at the end of the reach	Normal elbow extension (extent of elbow extension at the end of reach to grasp is consistent with the less-affected side)	Normal hand opening and closure to perform the task relative to the other side	The task is completed on the first attempt, but performance is not consistent with the less-affected side (arm bumps table during reach; abnormal timing or sequence of movement; tremor; dysmetria; arc-shaped trajectory)
4						The task can be done easily on the first attempt; the timing and sequence of trunk/shoulder/elbow/hand movements are consistent with the less-affected side

Figure 1. Final WMFT-RPSS Rating Criteria.

The table lists the final WMFT-RPSS scoring criteria that were used to rate *Can* and *Box* task performance (See the Expanded Methods, Appendix 1 for detailed description of modifications made to the original RPSS criteria). Polygon plots illustrate individual performance on the six *Can* and five *Box* WMFT-RPSS elements relative to the maximum possible score for the *Can* and *Box* tasks. Each line of the polygon plot represents a numeric value (0-4) that corresponds with the element score. Max possible score is represented by the black polygon and one participant's score is represented by the red polygon. In these examples, a score of 1 was assigned for each element, representing a total score of 6 out of 19 possible points for the Pick Up Can Task and 5 out of 16 possible points for the Hand to Box Task.

Figure 2. Association between Task Movement Rate (X-axis) and WMFT-RPSS Total Score (Y-axis) for the Can (Figure 2a) and Box tasks (Figure 2b).

Participant element scores are represented as polygon plots, as shown in the inset in the lower right corner. The black line represents the maximum scores possible for each element (4 for Task Completion, 3 for all other elements) and the red line denotes the individual's score. The polygon plots illustrate inter-individual WMFT-Reaching Performance Scale for Stroke variability in movement quality and compensatory movements among individuals with similar total WMFT-RPSS scores and movement times. There was a significant correlation between movement rate and total WMFT-RPSS score ($p < 0.0001$). The yellow and blue highlighted boxes identify cases in which participants have the same movement rate but different WMFT-RPSS total scores (2A), and participants with the same WMFT-RPSS total score but different element score patterns and movement rates (2B).