Interrater and Intrarater Reliability of the Wheelchair Skills Test Version 4.2 for Power Wheelchair Users

Emma M Smith[123], Kimberly G Low[4], William C Miller[234]

[1]Graduate Program in Rehabilitation Sciences, University of British Columbia, Vancouver, Canada; [2]Rehabilitation Research Program, GF Strong Rehabilitation Centre, Vancouver, Canada; [3]International Collaboration on Repair Discoveries, Vancouver, Canada; [4]Department of Occupational Science and Occupational Therapy, University of British Columbia, Vancouver, Canada

Published: January 23, 2017

Abstract

Purpose: To estimate the interrater and intrarater reliability of the Wheelchair Skills Test Version 4.2 for powered wheelchairs operated by adult users.

Materials and Methods: Cohort study with a convenience sample of occupational therapists (n = 10). For the main outcome measure, participants viewed and scored eight videos of adult power wheelchair users completing the 30 skills of the Wheelchair Skills Test Version 4.2 on two occasions, a minimum of two weeks apart. Using these scores, we calculated intraclass correlation coefficients to estimate interrater and intrarater reliability.

Results: The interrater reliability intraclass correlation coefficient was 0.940 (95%CI0.862-.985). Intrarater reliability intraclass correlation coefficients ranged from 0.923-0.998.

Conclusions: The Wheelchair Skills Test Version 4.2 has excellent interrater and intrarater reliability and is a reliable tool for use in clinical and research practice to evaluate a power wheelchair user's skill capacity.

Key Words: Outcome assessment (healthcare); Rehabilitation, Reproducibility of Results, Wheelchairs, Clinician

Word Count: 3974

Introduction

Power wheelchairs provide opportunities for increased independence for individuals with mobility disabilities, as well as improved quality of life, well-being, and self-esteem by facilitating participation in meaningful activities.**[1–4]** In the United States, there are over 3.6 million non-institutionalized users of wheelchairs.**[5]** In Canada, there were approximately 197,560 manual and 42,360 power wheelchair users living in the community in 2012.**[6]** The use of power wheelchairs is expected to continue to increase due to the aging population, growing incidence of disability, and evidence supporting the benefits of power mobility devices.**[3,7]**

Despite the benefits of power wheelchairs and their increasing use, barriers such as the natural environment and transportation can limit participation in desired occupations and community integration.**[1,3,8]** Commonly identified barriers to participation include narrow aisles, uneven ground, and tight spaces that require the wheelchair user to maneuver backwards.**[3,9]** There are also concerns from the wheelchair users themselves, as well as from friends, family, and health professionals, regarding the power wheelchair user's safety and risk for accidents.**[2,3,10]**

There is a role for rehabilitation therapists to provide assessment and training of wheelchair skills to help power wheelchair users overcome these challenges. Wheelchair skills assessments are used in practice to help identify areas of difficulty, guide clinical intervention, and as outcome measures to monitor a client's progress in skills training.**[11–13]** Training can be provided for these skills to increase confidence and remove barriers. Wheelchair training programs have been shown in studies with powered wheelchair users to improve wheelchair skills and confidence.**[14,15]**

There are limited outcome measures available to assess power wheelchair skills. For those which do exist, there is a need for further evaluation of the measurement properties of these tools.**[16–19]** The Wheelchair Skills Test (WST) is an objective standardized assessment and is available free online.[**[20]**] There are different versions and iterations of the WST including those for manual wheelchairs (WST-M) or power wheelchairs (WST-P) either operated by the user or a caregiver, as well as a self-report questionnaire version. Version 4.2 of the WST-P (WST-P 4.2) has 30 wheelchair skills and is scored on a 3-point scale (0-2) to reflect the wheelchair user's capacity to complete each skill

Although used in recent research,**[14,15]** there are no published measurement properties for any of the WST-P versions except the Wheelchair Skills Test Questionnaire (WST-Q 4.1) (for powered wheelchairs), which found support for reliability, validity, and responsiveness.**[19,21]** Although the questionnaire version has excellent measurement properties, the objective version is often used in clinical and research practice to assess baseline and post-intervention skill development. Establishing the reliability of this measure will strengthen the rationale for its use in research, as well as its use as a clinical measure in practice. Our objective is to estimate the interrater and intrarater reliability of the WST-P 4.2 for adult power wheelchair users operating their own chairs.

Materials and methods

Participants

We used a convenience sample of 10 raters. Raters were adult, English speaking occupational therapists with clinical and/or research experience working with people who use

power wheelchairs. Informed consent was obtained from each participant. This study was approved by the Behavioural Research Ethics Board of [XXX].

Wheelchair Skills Test for Powered Wheelchair Users Version 4.2

The Wheelchair Skills Test for Powered Wheelchair Users is an objective measure of wheelchair skill capacity for 30 powered wheelchair driving skills. Capacity is defined as what the wheelchair user *can* do and is not reflective of whether or not the individual *does* do it on a regular basis (performance).**[20]** The assignment of scores for capacity is according to published criteria based on the raters' judgement of how well and safely each skill is performed. If an individual completes a skill "independently and safely … without any difficulty", they are given a score of 2. If they are unable to complete the skill within the defined parameters, or decline to attempt, they are given a score of 0. A score of 1 is given if the tester feels there was "difficulty worthy of note" completing the skill (e.g. requiring additional time), or if the individual was unsafe while completing the skill (but did not require trainer intervention). If the skill is not possible (e.g. the wheelchair does not have the part required to complete the skill), the tester provides a score of Not Possible (NP). The notation of Tester Error (TE) is used if there was an error in conducting the test for that specific skill. The total score for the test is derived by adding the scores for all skills, dividing by the total possible score ((30 items – n items scored TE or NP) x2), and multiplying by 100 for a percentage score.

Videos

We used video recordings of test subjects performing the WST-P. We obtained eight videos, four from the Wheelchair Skills Program website **[22]**and four from a previous research study conducted by one of the investigators.**[23]** Consent to use the videos for research purposes was obtained from the individuals in the videos. Administration of the WST-P in the videos was

performed according to the WST-P 4.2 Manual by individuals independent to the raters and authors. The videos featured test subjects who were adult power wheelchair users with a variety of abilities ranging from full upper extremity use to high levels of upper extremity tone or limited fine motor control. All test subjects featured in the videos used a standard or modified joystick drive and were driving either a mid (5) or rear (3) wheel drive wheelchair. Test subjects were encouraged to attempt each skill but had the right to decline any that they felt they were unable to complete safely. While the WST-P takes approximately 30 minutes to administer, the videos were edited with each skill titled and unnecessary footage removed (e.g. transitioning between skills) to reduce participant burden removed. Each edited video was between 8 and 11 minutes in length.

Data collection

Raters were provided with a brief training period at the outset of the study. This training consisted of a review of available materials, as well as specific training on scoring criteria for the WST-P 4.2. This was reviewed at the start of each rating session. The WST-P 4.2 manual was provided for reference to each rater. Each of the raters independently viewed the videos on tablets that had videos pre-loaded in a randomized order to minimize potential order effect. Raters were instructed to score the test subjects based on the WST-P 4.2 criteria and to record their scores on the WST-P 4.2 scoring forms.**[17]** Raters were also given the option to write comments for each skill. Participants were instructed to score a "0" for skills that were refused and a "Testing Error (TE)" for skills that were not shown on the video or would require the rater to make assumptions due to inadequate footage. Participants were also instructed to score a "Not Possible (NP)" for any skill which required a part on the wheelchair which was not available (e.g. presence of tilt/recline functions). Each rater viewed all eight videos on two separate

occasions (R1, R2) a minimum of two weeks apart. Total percentage scores were calculated following WST-P 4.2 scoring procedures.**[17]**

Data analysis

SPSS Statistics 24 [IBM] was used to conduct all statistical analysis. Means and standard deviations were calculated for all quantitative demographic data. Interclass Correlation Coefficients (ICCs) were calculated to estimate the reliability of the total percentage scores of the WST-P 4.2. Standard Error of Measurement (SEM) and Smallest Real Difference (SRD) were calculated for each of the ten raters to address responsiveness, with Bland Altman plots constructed to identify potential areas of systematic bias in ratings**[24,25]**. Mean percent agreement across all videos for each skill was also calculated. Interrater reliability (ICC [2,10]) was calculated by comparing all R1 total percentage scores from the 10 raters. Intrarater reliability (ICC [2,1]) was calculated by comparing the R1 and R2 total percentage scores for each rater individually. ICCs ≥ .75 were interpreted as excellent, ICCs > .40 as good to moderate and ICCs < .40 as poor reliability.**[26,27]** Percent agreement was calculated by dividing the number of R1 scores in agreement by the total number of scores for each skill in each video. Mean percent agreement (and standard deviation) was calculated by averaging the percent agreement across all videos for each skill, as indicated in the equation below.

$$mean\ \%\ agreement\ for\ skill\ y = \frac{(P(a) skill\ y\ video\ 1) + (P(a) skill\ y\ video\ 2) + \cdots + (P(a) skill\ y\ video\ 8)}{total\ \#\ videos\ (n = 8)}\ x100$$

P(a) = percent agreement

We established an a priori cut off of 90% for percent agreement, to ensure agreement was not due to chance, as suggested by Neuendorf **[28]**

Results

Participant demographics

Demographics for raters are summarized in table 1. Experience working with clients using power wheelchairs and providing power mobility skills training varied among raters. There was representation from both clinical and research practice, as well as a variety of practice areas. Two raters had no previous training with the WST.

[Insert table 1 here]

Reliability

The ICC results for interrater and intrarater reliability are shown in table 2. The interrater and intrarater ICC values were all in the excellent reliability range. The interrater ICC (2, 10) was .940 (n=10, 95% confidence interval [CI], .862-.985). The intrarater ICCs (2, 1) for each rater ranged from .923-.998. SEM ranged from 0.584 to 2.814 across ten raters, while SRD for each rater ranged from 2.118 to 4.650. The Bland-Altman plots showed an even distribution of T1-T2 score difference across the mean difference line, with only one outlier across ten raters and eight test subjects (figure 1). Mean percent agreement and standard deviation for each skill is shown in table 3. Nine skills had 100% agreement across all raters and all videos. Three skills had percent agreement below 90%.

[Insert table 2 here]

[Insert table 3 here]

Discussion

This was the first study to estimate the reliability of the WST-P 4.2 and found both excellent interrater and intrarater reliability, showing substantial agreement between and within raters. Our results for intrarater reliability were similar to the findings reported for the WST-M 4.1 performance scores and WST-M 2.4.**[29,30]** Our interrater reliability results were higher than

those reported for the WST-M 4.1 performance scores (ICC=.855) and slightly lower than those reported for the WST-M 2.4 (ICC=.959).**[29,30]** Excellent interrater and intrarater reliability suggests scores will remain relatively stable across raters, and within the same rater; therefore, scores may be considered reliable in so far as they measure the construct consistently. This means the measure is likely to provide consistent measurement for clinical or research purposes.

Differences in interrater reliability scores between the Wheelchair Skills Test for Manual Wheelchair Users (WST-M) and the WST-P may be explained in part by sample size; however, fundamental differences in the skills required for manual and powered wheelchair users also play a role. With respect to individual wheelchair skills, there may be less subjectivity in the scoring of the WST-P skills than for the WST-M, which could explain our higher interrater ICCs. For example, many of the WST-P skills have obvious pass/fail distinctions (e.g. turning on and off controller, swing away the controller, and navigating drive modes, etc.). Given that two of the raters did not have previous experience with the WST-P, it is interesting to note that ICCs remained high despite minimal training. This may be explained by the more obvious pass/fail distinctions for many of the skills in the WST-P as compared to the WST-M. In addition, the differences in the scoring criteria introduced between the WST-M 2.4 and the WST-P 4.2, with a change from a 2-point scale to 3-point scale, and the combining of performance and safety scores (now a single capacity score), may influence differences in the rater reliability.

All values for the Smallest Real Difference were below 5 points, which represents a change in 1.5 skills on the WST-P to identify a difference across two administrations of the measure. The range in SRD values may be influenced by the experience of the raters, however it is interesting to note that these values were all low, demonstrating excellent responsiveness, regardless of rater experience. The distribution of the Bland Altman plots show no systematic

bias in ratings, although it is difficult to assess whether additional bias may be present at lower scores, as all of the participants' scores on the WST were high.

Differences in scoring of individual items may contribute to the variation found in the scoring of the WST-P. Three of the items (maneuvers sideways, transfers to and from a bench, and ascends a low curb) scored below 90% in mean percent agreement across raters and videos. There are various potential sources of disagreement in scoring. Disagreement may result from differences in the judgement of safety, as each rater applies their own clinical judgement to the criteria. For example, it is relatively straightforward to assess whether a client has successfully turned the wheelchair on and off, while assessing safety in completing a transfer between the wheelchair and a level surface allows for subjective assessment by the rater. Clearer scoring criteria may result in higher reliability as less interpretation is needed by the rater.

A strength of our study is our sample size of 10 raters, whereas the WST-M 4.1 had four and the WST-M 2.4 had two for estimating interrater reliability.**[29,30]** Our sample exceeds the amount determined from the WST-M 4.1 study, which identified a sample size of nine videos and four raters was necessary to achieve a power of 0.80 with an α level of 0.05 and interrater ICC values of 0.950 and 0.959.**[30]** Although we used a convenience sample, and the results may not necessarily generalize to the entire population of potential raters, we captured heterogeneity in raters' experiences working with individuals who use power wheelchairs, providing power wheelchair skills training, and practice areas. The video participants also had a variety of abilities, impairments, and drive control systems.

The strong interrater and intrarater reliability supports the use of the WST-P 4.2 in research and clinical practice. We have demonstrated that a group of occupational therapists with a variety of experiences can provide consistent scores for the WST-P 4.2 between each other and

within the same rater. It is important; however, that wheelchair skills testing be used to contribute to an overall assessment and not solely relied upon to assess capability for wheelchair driving. Although the WST-P 4.2 assesses specific skill capacity, understanding the context in which a wheelchair user wants and needs to operate their chair in as well, as the activities they want and need to participate in, is also important. Wheelchair skills testing can be used to help identify potential environmental barriers as well as inform areas where training can be provided to help reduce these barriers and increase a wheelchair user's skill level and confidence.

Study limitations

There were limitations in our study, the most obvious being the use of videos rather than live scoring. Three videos were missing a skill, and a few skills were difficult to view due to the footage filming angles. This may have prevented raters from viewing the necessary footage/information, including lines drawn on the floor which are used to gauge performance, and the position of the caster wheels in relation to these lines and other obstacles which inform accurate scoring. In regular use of the WST, test administration and scoring would likely be completed live by the same individual. Future studies opting to use videos should ensure that the footage allows all pertinent information to be seen.

With respect to intrarater reliability, the time of two weeks between R1 and R2 may not have been enough to prevent raters from recalling their previous ratings. Although raters viewed eight videos totalling 240 skills and approximately 2.5 hours each session, it is possible that memory could be a factor supporting high intrarater ICCs. Future studies could use a longer time period between R1 and R2. There is also an opportunity for future research to assess the test-retest reliability and validity of the WST-P 4.2 to provide further support for use of this measure. Additionally, the use of a volunteer sample of occupational therapist raters may not be

representative of the entire population that would potentially use the WST-P 4.2. In particular, two raters were novice users of the WST. This may have resulted in difficulties in scoring, which were not mitigated by the training provided. However, our findings suggest that the measure remains reliable regardless of the experience of the rater.

Conclusions

This was the first study to estimate the reliability of the WST-P 4.2. The WST-P 4.2 had excellent interrater and intrarater reliability. There is support for the use of the WST-P 4.2 as a reliable tool for use in evaluating an adult power wheelchair user's skill capacity. Further study of the test-retest reliability and validity of the WST-P 4.2 would be beneficial to establish the measure's reliability when used in pre and post-test comparison.

References

1.. Brandt Å, Iwarsson S, Ståhle A. Older people's use of powered wheelchairs for activity and participation. J Rehabil Med. 2004 Mar 1;36(2):70–7.

2. Edwards K, McCluskey A. A survey of adult power wheelchair and scooter users. Disabil Rehabil Assist Technol. 2010;5(6):411–9.

3. Fomiatti R, Richmond J, Moir L, Millsteed J. A Systematic Review of the Impact of Powered Mobility Devices on Older Adults' Activity Engagement. Phys Occup Ther Geriatr [Internet]. 2013 Dec [cited 2014 Jan 30];31(4):297–309. Available from: http://informahealthcare.com/doi/abs/10.3109/02703181.2013.846451

4. Pettersson I, Törnquist K, Ahlström G. The effect of an outdoor powered wheelchair on activity and participation in users with stroke. Disabil Rehabil Assist Technol [Internet]. 2006 Jan [cited 2013 Dec 9];1(4):235–43. Available from: http://informahealthcare.com/doi/abs/10.1080/17483100600757841

5. Brault, MW, Americans with disabilities: 2010. Current Population Reports, P70-131, U.S. Census Bureau, Washington, DC. [Internet]. 2012:70-131. [cited 2016 Nov 1] Available from http://www.census.gov/prod/2012pubs/p70-131.pdf

6. Smith E, Giesbrecht E, Mortenson W, Miller W. The Prevalence of Wheelchair and Scooter Use Among Community-Dwelling Canadins. Phys Ther. 2016;Epub ahead.

7. World Health Organization. Health and Ageing: A discussion paper [Internet]. 2001. [cited 2016 Nov 1] Available from: http://whqlibdoc.who.int/hq/2001/WHO_NHM_HPS_01.1.pdf

8. Lysack C, Komanecky M, Kabel A, Cross K, Neufeld S. Environmental Factors and Their Role in Community Integration after Spinal Cord Injury. Can J Occup Ther [Internet]. 2007;74(5 Suppl):243–54. Available from: http://www.researchgate.net/publication/5995554_Environmental_factors_and_their_role_in_community_integration_after_spinal_cord_injury

9. Torkia C, Reid D, Korner-Bitensky N, Kairy D, Rushton PW, Demers L, et al. Power wheelchair driving challenges in the community: a users' perspective. Disabil Rehabil Assist Technol [Internet]. 2015;10(3):211–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24640944

10. Mortenson WB, Miller WC, Boily J, Steele B, Odell L, Crawford EM, et al. Perceptions of power mobility use and safety within residential facilities. Can J Occup Ther [Internet]. Clinical Practice Leader (OT), Vancouver Coastal Health Authority, Rehabilitation Services, University of British Columbia Hospital, 2211 Westbrook Mall, Vancouver, BC V6T 2B5; 2005 Jun 1 [cited 2013 Dec 9];72(3):142–52. Available from: http://cjo.sagepub.com/lookup/doi/10.1177/000841740507200302

11. Best KL, Kirby RL, Smith C, MacLeod DA. Wheelchair skills training for community-based manual wheelchair users: A randomized controlled trial. Arch Phys Med Rehabil. 2005;86(12):2316–23.

12. Kirby RL, Swuste J, Dupuis DJ, MacLeod DA, Monroe R. The Wheelchair Skills Test: a pilot study of a new outcome measure. Arch Phys Med Rehabil [Internet]. Queen Elizabeth II Health Sciences Centre, Rehabilitation Centre, 1341 Summer St, Halifax, NS B3H 4K4, Canada; 2002;83(1):10–8. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=2002062551&site=ehost-live

13. Routhier F, Kirby RL, Demers L, Depa M, Thompson K. Efficacy and retention of the french-canadian version of the wheelchair skills training program for manual wheelchair users: A randomized controlled trial. Arch Phys Med Rehabil. 2012;93(6):940–8.

14. Mountain AD, Smith C, Kirby RL. Are wheelchair-skills assessment and training relevant for long-standing wheelchair users? Two case reports. Disabil Rehabil Assist Technol [Internet]. 2010 May [cited 2013 Dec 9];5(3):230–3. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/20131977

15. Kirby R, Miller W, Routhier F, Demers L, Mihailidis A, Polgar J, et al. Effectiveness of a Wheelchair Skills Training Program for Powered Wheelchair Users: A Randomized Controlled Trial. Arch Phys Med Rehabil. 2015;Epub ahead.

16. Dawson D, Chan R, Kaiserman E, E DD; CR; K. Development of the Power-mobility Indoor Driving Assessment for residents of long-term care facilities: a preliminary report. Can J Occup Ther [Internet]. Dept Occupational Therapy, Univ Toronto, Toronto Ontario, Canada; 1994 Dec 1 [cited 2014 Jan 16];61(5):269–76. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=1995002585&site=ehost-live

17. Kirby RL, Smith C, Parker K, McAllister M, Boyce J, Rushton P, et al. Wheelchair Skills Program Manual [Internet]. Halifax, NS; 2016. Available from: www.wheelchairskillsprogram.ca/eng/manual.php

18. Mills TL, Holm MB, Schmeler M. Test-retest reliability and cross validation of the functioning everyday with a wheelchair instrument. Assist Technol. 2007;19(2):61–77.

19. Mortenson W, Miller W, Auger C. Issues for the selection of wheelchair-specific activity and participation outcome measures: A review. Arch Phys Med Rehabil. 2008;89(6):1177–86.

20. Kirby RL, Smith C, Parker K, McAllister M, Boyce J, Rushton PW, et al. Wheelchair Skills Program Manual. 2016; Available from: www.wheelchairskillsprogram.ca/eng/manual.php

21. Rushton PW, Kirby L, Routhier F, Smith C. Measurement properties of the Wheelchair Skills Test - Questionnaire for powered wheelchair users. Disabil Rehabil Assist Technol. 2014;Epub ahead.

22. Kirby RL. Wheelchair Skills Program [Internet]. 2016 [cited 2016 Nov 3]. Available from: http://www.wheelchairskillsprogram.ca

23. Rushton PW, Kirby R, Routhier F, Smith C. Measurement properties of the Wheelchair Skills Test Questionnaire for powered wheelchair users. Disabil Rehabil Assist Technol. 2014;11(5):400–6.

24. Weir JP. Quantifiying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.

25. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60.

26. Andresen EM. Criteria for assessing the tools of disability outcomes research. Archives of

Physical Medicine and Rehabilitation. 2000.

27. Fleiss J, Levin B, Cho Paik M. Statistical Methods for Rates and Proportions. John Wiley & Sons. 2003. 800 p.

28. Neuendorf KA. The Content Analysis Guidebook. 2001.

29. Kirby RL, Dupuis DJ, MacPhee AH, Coolen AL, Smith C, Best KL, et al. The Wheelchair Skills Test (version 2.4): measurement properties. Arch Phys Med Rehabil [Internet]. Division of Physical Medicine and Rehabilitation, Department of Medicine, Dalhousie University, Halifax, NS, Canada; 2004 May [cited 2013 Dec 16];85(5):794–804. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0003999303009754

30. Lindquist NJ, Loudon PE, Magis TF, Rispin JE, Kirby RL, Manns PJ. Reliability of the Performance and Safety Scores of the Wheelchair Skills Test Version 4.1 for Manual Wheelchair Users. Arch Phys Med Rehabil [Internet]. Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Alberta; 2010 Nov;91(11):1752–7. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=2010845871&site=ehost-live

Figure captions

figure 1. Bland-Altman plots comparing T1-T2 scores to T1&T2 mean score.