Reviewer ID: Jeff Tan, Kyle Diab, Matthew Querée, Sharon Jang, Gurmaan Gill			
Type of Outcome Measure: Wheelchair Skills Test (WST, WST-Q)			**Total articles:** 9
Author ID Year	**Study Design**	**Setting**	**Population (sample size, age) and Group**
Kirby et al. 2002	Methodological pilot study with within-subject comparisons	Rehabilitation Center	N=24 wheelchair users (16M, 8F) Age: 59±19 Note: Only 12% SCI 11 amputees, 4 stroke, 3 musculoskeletal disorders, 3 with SCI, 3 with neuromuscular disorders Inpatients and Outpatients
Kirby et al. 2004	Cohort Study to assess measurement properties of the WST 2.4	Rehabilitation Center	N=298 (140M, 158F); 129 able bodied, 62 amputees, 20 musculoskeletal conditions, 34 SCI, 52 stroke and acquired brain injury. Age for SCI: 41.9 ± 20.6 Note: Only 11% SCI
Pradon et al. 2012	Methodological study	Not specified	N=40 (30M, 10F) Mean age: 36.9±11.2y Mean (range) 79.8 (1-360) months in rehabilitation SCI patients Tetraplegic (C6-C7): 7 High paraplegic (T1-T9): 15 Low paraplegic (T10-L3): 18
Lemay et al. 2011	Cross-sectional study	Institut de readaptation en deficience physique de Quebec (IRDPQ) (n=29) and the Toronto Rehabilitation Institute (TRI) (n=25)	N=54 (41M, 13F) Mean (SD) age: 46.7 (12.8) Manual wheelchair experience: 16.0 (11.2) years SCI patients Tetraplegic (C4-C8): 14 (25.9%) High paraplegia (T1-T6): 10 (18.5%) Low paraplegia (T7-L2): 30 (55.6%)
Hosseini et al. 2012	Cross-sectional multisite study	Six spinal Cord Injury Model Systems (SCIMS) centers	N = 214; Mean age = 38.8 (12) years; Mean time since injury = 11.7 (11) years (range = 0.9-50 years); Level of injury = C3-L5, 72% with paraplegia, 28% with tetraplegia
Rushton et al. 2016	Test-retest study assessing measurement properties of WST-Q for power users	Community-dwelling experienced power wheelchair users (Multi-city)	N = 72 (36M); Mean age = 60.7 (7.3) years (range 50-77). Note: Sample only 19% SCI
Lindquist et al. 2010	Test-retest design with 1-2 weeks in between testing.	Community-dwelling manual wheelchair users - University research clinic	N=11 (9M) Mean (SD) Age: 41.2±16.2 9 SCI, 1 stroke, 1 Arteriovenous malformation No info on SCI injury types Mean (range) overall wheelchair experience 9.7(1-37) years
Mountain et al.	Within participant	Community-dwelling	Only 20% SCI

2004	comparisons – WST and WST-Q for Manual wheelchair users	manual wheelchair users Rehabilitation Center	Participants: N=20 manual wheelchair users (12M, 8F), Age: 64.3±13.6, 7 Amputees, 4 musculoskeletal disorders, 4 SCI, 5 stroke and acquired brain injury
Kirby et al. 2016	Cross-sectional survey	Four spinal cord injury model systems centers	N=117, 100 male Median age (IQR): 39 (29.3-49.8) Median BMI (IQR): 25.8 (21.8-30.5) 90 paraplegia, 26 tetraplegia
Passuni et al. 2018	Cross-sectional survey	Rehabilitation unit, FLENI Institute, Buenos Aires, Argentina	N=11, 10 male Mean age (SD): 29.81 (12.18) years 11 wheelchair users 10 cannot walk, 1 can walk1

1. RELIABILITY

Author ID	Internal Consistency	Test-retest, Inter-rater, Intra-rater
Passuni et al. 2018		Inter-rater reliability: ICC = 0.998
Kirby et al. 2002	No data available	ρ = 0.65 (P=.001) Intra-rater: ρ = 0.96 (P<.001) Inter-rater: ρ = 0.95 (P<.001)
Kirby et al. 2004	No data available	Interrater – ICC = 0.968 Intrarater – ICC = 0.959 Test-Retest – ICC = 0.904
Pradon et al. 2012	No data available	The intraclass correlation coefficients (ICC) evaluating test-retest reliability for V_{max}, V_{spont}, and S_{time} were 0.94, 0.84, and 0.88, respectively. The ICC evaluating inter-rater reliability for V_{max}, V_{spont}, and S_{time} were 0.92, 0.92, and 0.95, respectively.
Rushton et al. 2016	Cronbach's alpha coefficient was 0.90.	The one-month test–retest reliability $ICC_{1,1}$ was 0.78 with a confidence interval of 0.68–0.86.
Lindquist et al. 2010		Interrater – ICC=0.855 P<0.001 Intrarater – ICC = 0.950 P<0.001 Test-retest – ICC = 0.901 P<0.001

2. VALIDITY

Author ID	Validity
Kirby et al. 2016	Correlation of WST capacity with WST-Q capacity: ρ=0.55
Kirby et al. 2002	The participants' occupational therapists unanimously endorsed 30 (91%) of the 33 WST skills. The therapists felt that 13 of the 21 subjects improved and 8 had not changed between tests 1 and 2. The paired t tests between the subjects' total raw WST scores on test 1 and 2 showed a mean improvement of 2.1± 5.7 (P<.05). Improved: 3.2±5.9 No change: 0.5±5.2 (p=.15)

	Mean total WST scores 39.8±6.7 (n.s.) Mean changes WST & VAS: ρ = 0.45 (P=.025) The changes in total WST and VAS between WSTs 1 and 2 were statistically, but not clinically, significant. There is a slightly negative relationship between total scores on WST 1 and age. There is a significant positive relationship between total WST scores and duration of wheelchair use. <u>Rasch Analysis</u> From the 20 skills for which there were sufficient data, by sequential fitting of the model, 6 skills were dropped, because they added nothing to the model, were not in the same dimension as the other skills, or did not correlate well with them. The remaining 14 skills compromised a unidimensional test of wheelchair ability. <u>WST score & occupational therapists global assessment of manual wheelchair skills (Visual Analog Scale (VAS))</u> WST 1 & VAS: ρ = 0.40 (P=.01) WST 2 & VAS: ρ = 0.54 (P=.008)
Kirby et al. 2004	Construct Validity - Moderate negative correlation with age (r=-0.434, P<0.001) - Multiple regression shows age is a significant factor (P<.001) - Multiple regression shows gender is a significant factor (P<0.001) (Men = 69.2%±13.2%, Women = 67.8%±14.5%) Regarding construct validity, there was a slightly negative Pearson correlation between total WST score and age (-.434). Gender was identified as a significant factor on multiple regression analysis (P<.001). Wheelchair users with more than 21 days of experience scored higher than those with less experience (65.0% vs 59.6%; P=.01). Participants with stroke and related disorders had a mean score (55.0%+/-13.9%) that was significantly lower than those in other diagnostic categories (P<.05). Participants using conventional wheelchairs had lower scores than those in lightweight ones (66.4% vs 75.1%; P<.001). Regarding concurrent validity, Spearman rank correlations between total WST scores and the 1) subjective evaluation of wheelchair users by their clinicians = 0.39; and 2) admission and discharge FIM instrument scores were 0.38 and 0.31.
Pradon et al. 2012	Spearmans' rank correlation coefficient between Wheelchair Skill Test score and the: - maximal velocity (V_{max}) = 0.72; - Spontaneous velocity (V_{spont}) = 0.57; - Slalom time (S_{time}) = -0.75 (P<.05 for all)
Lemay et al. 2011	Correlations between WST total performance score and: Wheeled distance per day: 0.36 (P<.01) Wheeled speed per day: 0.22 (ns – P>.05) Age: -0.32 (P<.05) Manual wheelchair experience: -0.15 (ns – P>.05) There was a significant correlation between WST total performance score and mobility (r=0.36, P<.01), which suggests that greater MWC skills are slightly associated with higher mobility at home and in the community in terms of daily wheeled distance. The correlation between WST total performance score and daily wheeled distance did not remain significant when controlled for age (partial r=0.26, P=.07). A moderate correlation between age and WST total performance score was also observed (r=-0.32, P<.05).
Hosseini et al. 2012	Predictive Criterion Validity WST version 4.1 score predicted CHART score, Satisfaction with Life Scale score, and self-perceived health status on multiple linear regression ($p < 0.05$)
Rushton et al. 2016	Concurrent Validity LSA (assisted life space) 2.0 (0–5) 0.47, p < 0.001 WheelCon-P (/100) 76.6 (13.5) 0.47, p < 0.001

	WST (/100) 80.1 (10.5) 0.65, $p < 0.001$
Mountain et al. 2004	Concurrent Validity - WST highly correlated with the WST-Q (r=0.91)
3. RESPONSIVENESS – no data available	
4. FLOOR/CEILING EFFECT – no data available	
5. INTERPRETABILITY	
Author ID	**Interpretability**
Kirby et al. 2002	WST 1 mean (SD) score: 36 (8) WST 2 mean(SD) score: 38 (10)
Pradon et al. 2012	The total group of participants obtained a mean WST score of 83.8% (SD 16.6%) (range 47.0–100.0%), a mean V_{max} of 6.91 km/h (SD 2.10) (range 3.6–11.7), a mean V_{spont} of 4.74 km/h (SD 1.21) (range 2.6–7.5) and a mean S_{time} of 60.4 s (SD 29.6) (range 25.0–153.0).
Lemay et al. 2011	Mean (SD) WST total performance score for different subgroups: All participants: 80.7±11.8 Tetraplegia: 72.1±7.9 High paraplegia: 82.8±9.1 Low paraplegia: 84.0±12.4 55.6% of participants scored over 80% (80% is empirically considered the cut-off for distinguishing people with advanced MWC skills (mainly skills required to control wheelies)) - 28.6% of the participants with tetraplegia (4 out of 14) reached or exceeded the 80% cut-off, suggesting advanced MWC skills.
Rushton et al. 2016	The SEM and SRD were 5.0 and 6.2 respectively. The SEM of 5.0 and the SRD of 6.2 represent the minimal change in WST-Q score that reflects a meaningful change beyond measurement error for a group and an individual respectively.