Measurement Reliability of Functional Tasks for Persons Who Self-Propel a Manual Wheelchair

Laura A. May, PhD, PT, Carla Butt, BScPT, Linda Minor, BScPT, Karen Kolbinson, BScPT, Kathy Tulloch, BScPT


Objective: To evaluate the reliability of 4 functional tasks relevant to wheelchair seating.

Design: Within-subject and between-rater comparisons.

Setting: Rehabilitation center in Canada.

Participants: Two separate convenience samples of 10 male wheelchair users.

Interventions: Not applicable.

Main Outcome Measures: The 4 functional tasks were timed forward wheeling, ramp ascent, forward vertical reach distance, and ramp descent, scored by an ordinal performance scale. To determine test-retest reliability, the participants performed each task twice on the same day. To determine inter-rater reliability, 5 experienced therapists independently scored each participant. The ramp descent task was replaced with a 1-stroke push distance task due to difficulties with the interpretation of the ordinal performance scale.

Results: Testing of all tasks was completed within 45 minutes, allowing for rest periods. There were no adverse incidents. One individual with C6 quadriplegia 4 months after spinal cord injury was unable to complete the ramp ascent. Estimates for test-retest reliability of all 4 functional tasks were excellent ($r=99$). Interrater reliability was calculated for all tasks except the 1-stroke push and found to be excellent (intraclass correlation coefficient $=99$).

Conclusions: The final 4 functional tasks are practical, safe, and reliable tests that may be used for clinical evaluation of wheelchair seating. Further research involving comparative assessments of wheelchair seating options is required to determine the discriminative ability of the tests.

Key Words: Outcome assessment (health care); Rehabilitation; Reproducibility of results; Wheelchairs.

© 2003 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

APPROPRIATE SELECTION of wheelchair and seating options is an important decision during rehabilitation and is influenced by the complex interaction of the environment and the person’s posture and mobility. Postural supports, including the back support and seat cushion, may influence sitting comfort, postural alignment, buttock pressure, and functional abilities. Although less costly than other options, the standard sling upholstery has distinct disadvantages resulting in a postural position that is often less than optimal. Various manufactured postural support options are available. Results of manufacturers’ product evaluations are often unavailable to the clinician or the consumer and scientific research regarding product benefits is not often published. Considering different effects and costs of seating components, clinicians and their clients must make informed choices, facilitated by accurate measurements. Because both the chair and the postural supports may affect functional tasks related to wheelchair propulsion, ascending and descending ramps, and reaching, it is important to include both static and dynamic evaluations. This report describes the development and reliability testing of such tasks.

Often clinicians rely on their product knowledge and clinical expertise when determining the most suitable product for the client requiring a manual wheelchair. With so many wheelchair models, seating systems, and accessories available, it can become difficult to keep up with the technology and be competent in making appropriate selections. Even if the choice is evident, the lack of scientific evidence to guide clinical judgment may be a problem. Objectively measuring the advantages and disadvantages related to seating posture and mobility often involves biomechanical or physiologic analysis, radiographs, or specifically designed laboratory instrumentation. The information regarding positioning or functional implications is valuable, but using similar tests for product comparison within a clinical setting may be difficult because of test complexity or equipment availability.

The dilemma lies in the lack of appropriate standardized clinical measures to evaluate functional abilities for manual wheelchair users within a relevant context. Although clients are given the opportunity to test drive different products, the objective evaluation by the clinician varies. A nonstandardized, program-specific checklist of wheelchair skill is generally used to confirm skill achievement rather than compare wheelchair and seating options. Standardized rehabilitation outcome measures are available to evaluate function or community integration. Unfortunately, these measures are not designed for wheelchair product comparison and are likely too general to be able to discriminate between options. Other measures cited in the literature offer profiles of specific functions simulating the daily activities encountered by power wheelchair users.

Efforts have been made to quantify wheelchair skills relevant to the adult spinal cord injury (SCI) population. A new instrument to test mobility of wheelchair-dependent persons with paraplegia by using 6 relevant skills has been shown to be reliable. Another recently developed instrument, the Wheelchair Skills Test (WST), is more comprehensive, including 50
skills, and is intended for a broader population of manual wheelchair users. Both tests have strengths and include potentially useful items. As tests to compare postural support options, however, they may have problems with their scoring scheme. The mobility test uses a 6-level criterion-referenced scale with set cut-scores for time, and the WST uses a dichotomous pass-fail score, although the original version used a 3-point ordinal scale. Actual time recordings may more sensitively detect differences in performance when using the postural support options.

Single-item tests may be valuable for comparing products. The reliability of the Modified Functional Reach Test has been evaluated as a test of sitting balance in individuals with SCI. Although this test is proposed as useful for comparing seating products, it is likely to be more useful for evaluating cushions or wheelchair seat angle than back supports, because the individual reaches forward horizontally and loses contact with the back support. Two ramp ascent activities testing propulsion speed and strength were developed for a study comparing back supports used by wheelchair dependent adults. To evaluate speed, the subjects propelled up a standard ramp as fast as possible, and the distance achieved with 1 push up the ramp was used to evaluate strength. Detailed testing procedures were not given nor was there any documentation of reliability of the tests.

It may not be practical to use existing measures that are either too broad in scope or too specific in the evaluation of postural support options. Therefore, a need exists for functional tests that may be used in a practice or research setting to permit objective comparison of postural support options. The purpose of the present study was to design and test the reliability of specific functional tasks that could be used by clinicians to quantify functional ability when persons who self-propel a manual wheelchair used different postural supports.

**METHODS**

**Identifying Functional Tasks for Testing**

Specific to the objective of postural support comparison, physical therapists on the SCI program at the Glenrose Rehabilitation Hospital in Edmonton, AB, all with more than 5 years of experience, developed a series of functional tasks in consultation with the clinical researcher and clients. Before developing a list of potential tasks, the literature was reviewed to determine if current single-item or multi-item functional tests would be appropriate for use or to provide guidance in selecting tasks. Set criteria were used to determine which tasks could be included. All tasks had to be (1) representative of daily activities encountered by wheelchair users, (2) potentially bi-mechanically able to be affected by the postural support provided by various wheelchair back options, (3) executable without relying on upper-extremity strength for stability (ie, "hooking" arm around the back cane), (4) executable with conventional equipment common to a clinical setting, and (5) sensitive enough in their scoring to detect differences in functional performance. Originally, 4 functional tasks were developed incorporating propulsion speed and strength and trunk stability. After preliminary testing, the task requiring descent of a ramp with a 180° turn was excluded. It was evident that the ordinal level rating was too subjective, which affected reliability. Furthermore, participants lacking adequate hand function were unable to complete the task without assistance. The final 4 tasks are briefly described later. All tests, except where indicated, were conducted indoors at the rehabilitation facility on a linoleum surface. Before beginning each task, the front casters of the wheelchair were so positioned that forward movement would not be affected by caster swivel.

**Timed forward wheeling.** Participants started from a standardized stationary position with anterior edges of the front casters positioned directly behind the start line. The start commands were consistent for all trials. Timing began from the first movement of the casters and ceased when the most posterior aspect of the rear wheels crossed the finish line. The wheeling distance of 23m (plus length of the wheelbase) was selected because it represents the average length of a crosswalk in a 4-lane intersection.

**Forward vertical reach.** The wheelchair was positioned parallel to the marking board, and participants held the measuring stick, independently or with cuff supports, using both hands (palms down) as seen in figure 1. The 1-m measuring stick had a felt pen attached to one end and a carpenter’s level secured in the middle. The tip of the felt pen was positioned within 2.54cm (1in) of the marking board. Participants began with forearms parallel to their thighs and raised the measuring stick upward. Participants were instructed to keep the measuring stick level and to avoid excessive backward arching. A dot was made with the felt pen on the marking board at the highest point reached. The distance in centimetres from the floor to the dot was measured. Bilateral reach was selected as a better test of the stability provided by the back support than unilateral reach.

**Ramp ascent (forward wheeling).** The participant was allowed a moving start of 4.5m on a level surface. The timing began when the front casters crossed the start line at the

---

**Fig 1. Participant showing the method used to measure forward vertical reach. Note the position of the carpenter’s level in the middle of the measuring stick.**
Participants
A total of 20 male wheelchair users volunteered for the study. At the time of the study, the inpatients and outpatients available, eligible, and willing to participate were men; we did not intend to exclude women. To be included in the study, participants needed to be medically stable, wheelchair dependent for more than 50% of the time, wheel independently, and able to give informed consent. Ten participants (avg age, 33.1y; range, 23–70y) took part in the test-retest and interrater reliability evaluation of the 4 original functional tasks. A separate group of 10 individuals (avg age, 29.9y; range, 19–57y) participated in test-retest reliability evaluation of the 1-stroke push task that was added after 1 of the original tasks, the ramp descent, was excluded. All participants had a neurologic condition affecting the spinal cord. Four of the first participant group and 6 of the second group had affected upper- and lower-extremity function, whereas the remainder had affected lower-extremity function only. All but 2 participants had been wheelchair dependent for less than 2 years. This study was approved by the Health Research Ethics Board (Panel B), a joint committee of the University of Alberta and the Capital Health Authority.

Procedures
For all testing, the participants used their current wheelchair and back support, and no adjustments to the wheelchair configuration were made during the test period. Two repeat trials were used to evaluate test-retest reliability. To determine interrater reliability, 5 physical therapists working in rehabilitation of persons with SCI and other neurologic conditions simultaneously recorded times for the 2 wheeling activities. Each therapist independently measured the distance marked for the reaching activity. To avoid upper-extremity fatigue during the testing period, participants were given a 5-minute rest between the 2 repeat trials and between different activities. This rest period allowed a minimum work to rest ratio of 2:1, which was considered sufficient to limit the effects of fatigue during testing. Before starting each test activity, the participant’s position in the wheelchair was checked by the same therapist to ensure optimal positioning for each individual. Participants were not repositioned during the test activity unless spasticity significantly altered positioning and continuation was not possible or the participant was at risk of falling or being injured. Participants were not informed of the test results. For the 1-stroke push task, which was added at a later date, only test-retest reliability was evaluated, and 1 therapist completed all the testing in the same format as described.

Analyses
All calculations were performed by using the SPSS, version 10.2. Descriptive statistics were used to present the participants’ scores for all the functional tasks. To determine test-retest reliability, Pearson product-moment correlation coefficients were calculated for the final 4 functional tasks and intraclass correlation coefficients (ICC 3,1) were used to evaluate the interrater reliability among the 5 raters for 3 of the original functional tasks.

RESULTS
The ranges and average scores for the final 4 functional tasks are in table 1. One individual with a C6 SCI, who was less than 4 months postinjury, was unable to complete the ramp ascent and thus the scores of only 9 participants were used to compute the reliability for this task. There were no other difficulties or adverse effects noted, and none of the participants complained...
of fatigue during testing. Reliability estimates for test-retest reliability of all 4 functional tasks was excellent ($r = .99$, $P < .001$). Figures 3 and 4 show the consistency with which the participants performed for both timed and distance-rated tasks. Interrater reliability was calculated for all tasks except the 1-stroke push and found to be excellent (ICC = .99, $P < .001$). The average scores (for all participants) of the 5 raters for the 3 tasks evaluated are in table 2.

### DISCUSSION

Although physical therapists are often involved in the evaluation and recommendation of wheelchair seating options, including back supports, objective measurement may not be available. Including objective measurement in the decision-making process could facilitate more informed chair selections. Multi-item gross motor evaluations include items that may not be relevant or may be difficult to use for product comparison, and the scoring schema may contribute to possible lack of sensitivity. When comparing products that provide postural support, test items that reflect the effects of the support’s performance may be particularly useful. The present study describes how 4 functional tasks relevant to wheelchair postural support comparison were developed, and showed their test-retest and interrater reliability. Although interrater reliability was not evaluated for the 1-stroke push task, the investigators judged that the distance measurement was similar to the reaching activity, and consequently the interrater reliability results could be extrapolated.

The functional tasks showed excellent reliability. The 4 tasks are simple and meaningful with respect to daily activity. Within a clinical setting, issues of time and equipment often limit the ability to obtain objective data. Approximately 30 minutes is required to complete 2 trials of the 4 functional tasks with adequate rest periods included. Furthermore, equipment needs are minimal and do not include anything that would not be available in a usual clinical setting. One disadvantage may be related to the difficulty of the ramp ascent activity. Early in a client’s rehabilitation, it may be necessary to delay performing the ramp ascent task or exclude it altogether if he/she is unable to complete it. Because reliability is reported for each task individually and the scores are not intended to be cumulative, single-test item results can be used. Another possible disadvantage relates to the method used for the 2 timed items. Although both tasks were reliable, reporting velocity would be difficult because the wheelbase was included in the total distance traveled.

With respect to wheelchair seating, different product brands intended for the same purpose may provide varying performance abilities for the same person. Considering the reliability results of the present study, it would be appropriate to use single-test item results, although there may be limitations associated with the interpretation of the performance informa-

<table>
<thead>
<tr>
<th>Trials</th>
<th>Forward Wheeling (s)</th>
<th>Forward Vertical Reach (cm)</th>
<th>Ramp Ascent (s)</th>
<th>One-Stroke Push (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.7</td>
<td>160.7</td>
<td>39.6</td>
<td>102.1</td>
</tr>
<tr>
<td>2</td>
<td>14.6</td>
<td>160.8</td>
<td>34.4</td>
<td>109.8</td>
</tr>
<tr>
<td>Average score</td>
<td>14.7</td>
<td>160.7</td>
<td>39.6</td>
<td>102.1</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>8.2</td>
<td>9.1</td>
<td>10.1</td>
<td>38.4</td>
</tr>
<tr>
<td>Minimum</td>
<td>6.9</td>
<td>147.4</td>
<td>3.2</td>
<td>45.7</td>
</tr>
<tr>
<td>Maximum</td>
<td>34.3</td>
<td>172.3</td>
<td>255.1</td>
<td>175.5</td>
</tr>
</tbody>
</table>

* One of the 9 participants who completed this task required multiple rests; that person’s score was considerably longer than other participants’, thus accounting for the large standard deviation.

![Fig 3](image1.png)  
**Fig 3.** Participants’ scores for the 2 trials used to determine test-retest reliability for the 1-stroke push task, which used a distance measurement ($r = .99$).

![Fig 4](image2.png)  
**Fig 4.** Participants’ scores for the 2 trials used to determine test-retest reliability for the forward wheeling task, which used a timed measurement ($r = .99$).
tion. The 4 tasks represent both static and dynamic activities, and it may be possible that some products that provide postural support contribute to improved performance in certain tasks but not all. For this reason we would recommend that all tasks be evaluated, if possible, and that results be compared within the context of the client’s goals and lifestyle. Although the parameters of each item will depend somewhat on the facility in which they are performed, reliability should be maintained given the objective scoring criteria of time and distance.

The scoring schema is important when considering the measurement purpose and the psychometric evaluation, which includes reliability. Use of interval-level measurement may enhance the reliability of the tasks. Timed and distance measurements reduce the possibility of subjective interpretation, which can influence other types of scoring. For example, the ordinal-level scoring of the ramp descent task was unreliable because of the inconsistent interpretation of its scoring categories. Although not tested in the present study, the timed and distance scoring of the 4 tasks may enhance the ability of these tests to be used for discriminative purposes. Therefore, it would be possible to document performance differences among persons with various skill levels, different diagnostic categories, or when using different back supports and seating options. The sensitivity of the incremental scoring could also facilitate the detection of changes over time when measurements are used for evaluative purposes. In a previous study evaluating wheelchair position and seating for persons with SCI, changes were shown when performance was measured by timed wheeling tasks similar to those described in the present study.

Although the scoring methods of time and distance would likely enhance reliability, some elements included in the tasks could be subject to inconsistent interpretation. Instruction to the participants, if not standardized, could produce unreliable results. The use of rests during the ramp ascent could also be open to variability between testers. Asking participants not to arch their backs during the reaching task may be enforced more aggressively by some therapists. Repositioning may also be a source of unreliability, although none of the participants in this study required repositioning. The issue of subjective interpretation of the tester influencing functional test results is not new and underscores the need for standardized instructions and training of all persons who will administer the test.

One question that remains to be answered refers to the validity of the tasks for the intended purpose of postural support product comparison. The tasks are assumed to have face validity considering the nature of the development, which included the input from experienced clinicians and wheelchair users. To be meaningful, the tasks should reflect what is important to both clinicians and clients with respect to functional performance. To be practical, the number of tasks must be kept to a minimum but still reflect diverse daily activities that could be affected by postural support. Objective scoring by means of timed or distance measurements enhance the interpretability of the results and communication between health professionals as well as between health professionals and clients.

**CONCLUSION**

This study showed the reliability of 4 functional tasks that may be used for clinical evaluation of wheelchair seating options, including postural supports. These functional tasks provide objective information, are simple to perform, and do not require excessive time or equipment. These results are primarily limited to wheelchair-dependent individuals with spinal cord dysfunction, although these tests may be useful with other patient populations who self-propel a manual wheelchair. Further research using these tasks in a comparative assessment of seating options would help to determine the discriminative ability of the tests. Evaluative studies would also help to determine the ability of these tests to detect changes over time. Additional research related to the validity of the 4 functional tasks is also required.

**References**


Supplier
a. SPSS Inc, 233 S Wacker Dr, 11th Fl, Chicago, IL 60606.