Department of Veterans Affairs

JRRD Volume 48, Number 2, 2011
Pages 115–124
Journal of Rehabilitation Research & Development

Telerehabilitation assessment using the Functioning Everyday with a Wheelchair-Capacity instrument

Richard M. Schein, PhD;[1*] Mark R. Schmeler, PhD, OTR/L, ATP;[1] Margo B. Holm, PhD, OTR/L;[2] Michael Pramuka, PhD;[1] Andi Saptono, MS;[3] David M. Brienza, PhD[1]

Departments of [1]Rehabilitation Sciences and Technology, [2]Occupational Therapy, and [3]Health Information Management, University of Pittsburgh, Pittsburgh, PA

Abstract—This study explored the interrater reliability between a generalist practitioner administering the Functioning Everyday with a Wheelchair-Capacity (FEW-C) in person (IP) and a remote expert practitioner observing via telerehabilitation (TR) from more than 100 miles away. Each of the 46 participants was simultaneously rated by both the IP and TR practitioner, who were masked to each other's results. The IP-TR raters demonstrated excellent interrater reliability, with an intraclass correlation coefficient of 0.91 for the total FEW-C and 0.96, 0.88, and 0.90 for the constructs of independence, safety, and quality, respectively. Internal consistency (Cronbach alpha) was 0.94 for the total FEW-C tool and 0.91, 0.83, and 0.82 for independence, safety, and quality, respectively, indicating good internal consistency without redundancy. Using TR and the FEW-C, an expert practitioner more than 100 miles away was able to accurately assess the functional mobility needs of clients being assessed for new wheeled mobility devices.

Key words: FEW-C, Functioning Everyday with a Wheelchair-Capacity, mobility limitation, physical function, rehabilitation, reliability, remote assessment, telehealth, telerehabilitation, videoconferencing, wheelchairs.

INTRODUCTION

The assistive technology community is accountable for quality of service as well as for identifying alternative and more efficient ways of delivering clinical services. For areas in which a shortage of rehabilitation professionals exists or specialists are unavailable, telerehabilitation (TR) is an option. TR can be defined as the application of telecommunication, remote sensing and operation, and computing technologies to the delivery of medical rehabilitation services at a distance. A growing amount of literature exists on the use of technology for remote assessment and intervention in medicine [1] and

Abbreviations: ATD = assistive technology device, CCMH = Charles Cole Memorial Hospital, CI = confidence interval, DRMC = DuBois Regional Medical Center, ERHC = Elk Regional Health Center, FEW-C = Functioning Everyday with a Wheelchair-Capacity, ICC = intraclass correlation coefficient, IP = in person *or* standards met–improvement possible, MMCHS = Meadville Medical Center Health System, MR = minor risk–no assist, NM = standards not met, PA = physical assist, PH = risk–potential harm, PM = standards partially met, RERC = Rehabilitation Engineering Research Center, SM = standards met, SP = safe practices, SR = severe risk–prevent harm, TR = telerehabilitation, VA = verbal assist, VSA = visual/setup assist, VISYTER = Versatile and Integrated System for Telerehabilitation, WMS = wheeled mobility and seating.

***Address all correspondence to Richard M. Schein, PhD; University of Pittsburgh, Department of Rehabilitation Sciences and Technology, 6425 Penn Avenue, Suite 401, Pittsburgh, PA 15206; 412-624-6429; fax: 412-624-6501. Email: rms35@pitt.edu**

DOI:10.1682/JRRD.2010.03.0039

rehabilitation [2–6]. Schmeler et al. recently reported on the state of the science of clinical and vocational assistive technology TR applications, outlining the current research, opportunities, and challenges [7].

Initial TR studies explored the use of low-bandwidth videoconferencing systems for consultations between clinicians [2], wheeled mobility and seating (WMS) evaluations [8], and orthotic assessments [9]. Several studies analyzed the use of TR in the field of WMS (i.e., manual/power wheelchairs and scooters). Cooper et al. compared the type of wheelchair the person actually used to the wheelchair recommended via in-person (IP) assessments versus TR assessments conducted with a videoconferencing system connected to a standard telephone line [10]. IP is defined as when a clinician is face-to-face with the participant. Clinicians who used TR demonstrated a high level of kappa agreement (0.76) in recommending the same basic type of wheelchair that subjects already owned, which also demonstrated a high level of kappa agreement (0.92) in consistency of wheelchair recommendation. Allegretti et al. reported interrater reliability for seven trunk alignment variables with the use of a videoconferencing system connected to a standard telephone line [11]. The results indicated that seven trunk alignment variables were only marginally better when measured by two therapists through an IP assessment than when measured IP by a therapy assistant with a therapist observing via TR.

In one pilot study, Dreyer et al. investigated the efficacy of using a less costly low-bandwidth telemedicine system to evaluate clients in rural areas by administering either the Kohlman Evaluation in Living Skills or the Canadian Occupational Performance Measure IP, while a second occupational therapist scored the same tool by the telemedicine link [12]. Comparison of responses revealed scoring differences in only one of four evaluations. Shafqat et al., using a high-speed videoconferencing system, administered the National Institutes of Health Stroke Scale to 20 patients with ischemic stroke with a stroke neurologist at bedside, while a second stroke neurologist observed via videoconferencing [13]. Based on weighted kappa coefficients, physicians achieved excellent agreement on four items, good agreement on six items, and poor agreement on two items. Palsbo et al. explored the equivalence of physical function assessments by physical therapists during IP and remote administrations of the European Stroke Scale and the Functional Reach Test [14]. Their conclusions were that when the remote physical therapist directed the patient, the therapists reported equivalent values in more than 83 percent of all European Stroke Scale components and more than 90 percent of the patients for the Functional Reach Test. In order for TR to continuously grow as an effective service delivery tool, we must establish the reliability of IP and TR methods for measuring performance. Other factors that must be considered with the growth of TR include development of guidelines and healthcare policies for appropriate clinical uses, advocacy for reimbursement (which is currently limited and inconsistent among private and public payers), and the effect on licensure laws that govern how professionals can practice.

The specific aims of this study were to (1) establish interrater reliability of the Functioning Everyday with a Wheelchair-Capacity (FEW-C) between an expert practitioner participating via TR and an IP generalist practitioner, and (2) determine the internal consistency of the FEW-C total and its three constructs: independence, safety, and quality.

METHODS

Study Participants

Potential participants were made aware of this research study and asked if they were interested in participating when making clinical appointments. All subjects were located at four remote sites, all of which were located at least 100 miles away from Pittsburgh. The study participants were recruited from DuBois Regional Medical Center (DRMC) in DuBois, Pennsylvania; Elk Regional Health Center (ERHC) in St. Marys, Pennsylvania; Charles Cole Memorial Hospital (CCMH) in Coudersport, Pennsylvania; and Meadville Medical Center Health System (MMCHS) in Meadville, Pennsylvania. Inclusion criteria were (1) 18 years of age or older and (2) use of a WMS device such as a manual or power wheelchair or scooter. The IP-TR reliability study was conducted with 46 participants: 25 manual wheelchair users, 18 power wheelchair users, and 3 scooter users. Since the participants had various primary diagnoses, the diagnosis categories were collapsed into the following: progressive (26.1%), spinal cord injury (10.9%), orthopedic (17.4%), cardiovascular (30.4%), and central nervous system (15.2%) (**Table 1**).

Outcome Instrument

Since no published criteria were available for selecting a measure appropriate for use with TR, a similar methodology to that used by Palsbo et al. was chosen for the selection of the FEW-C [14]. Criteria included (1) appropriate

Table 1.
Participant demographics.

Variable	Mean ± SD or %
Age	54.70 ± 15.46 (range 22–89)
Sex	
Female	63.0
Male	37.0
Race/Ethnicity	
Caucasian	89.1
Other	10.9
Primary Medical Condition	
Progressive	26.1
Spinal Cord Injury	10.9
Orthopedic	17.4
Cardiovascular	30.4
Central Nervous System	15.2
Type of Wheeled Mobility and Seating Device	
Manual Wheelchair	52.2
Scooter	6.5
Power Wheelchair	41.3

SD = standard deviation.

and relevant to people with mobility impairments who use WMS, (2) known psychometric properties (i.e., validity and reliability), (3) used in clinical practice, (4) performance-based and visual so that the practitioner can perform the measurement without touching the patient, and (5) administration time <45 minutes.

The FEW-C is a criterion-referenced, performance-based observation tool used by practitioners and researchers alike to measure functional outcomes of WMS users. The FEW-C focuses on the participant's ability to perform tasks or activities (e.g., mobility, reach, and transfer) in a controlled clinical or laboratory environment and takes between 45 minutes and 1 hour to administer.* The FEW-C consists of 10 criterion-referenced, performance-based tasks. Of the 10 tasks, several are strictly performance based: operate, reach and carry out tasks at different surface heights, transfers, personal care, indoor mobility, and outdoor mobility. Three tasks (comfort needs, health needs, and personal/public transportation) have both performance-based and self-report components because of the complexity (i.e., subjectivity and feasibility) associated with task measurement. Stability, durability, and dependability are self-report items that are subsequently measured during performance of all other tasks. For the purpose of this study, only the following tasks were recorded and scored: operate, reach, transfer, personal care, indoor mobility, comfort, and health needs. Because the videoconferencing system could not be moved outside, the tasks outdoor mobility and personal/public transportation were not scored. The performance-based items each yield three distinct category scores: independence, safety, and quality.

An example of a FEW-C data collection form is given in the **Figure**. For data collection, column 1 indicates the number and name of each FEW-C task. Column 2 indicates what type of mobility device (manual, power, or scooter) was used for task performance. If any assistive technology devices (ATDs) are used for task performance, the name(s) of the ATD is written on the provided space and the total number of devices used is indicated. At the lower end of column 2, each row contains the independence and quality data descriptions/examples, for which the independence, safety, and quality data are derived for task performance. Column 3 contains three hierarchical columns for the types of assistance (verbal assist [VA], visual/setup assist [VSA], and physical assist [PA]) that may be provided by the examiner during task performance. The type and number of assists is circled for each task. Column 4 contains four columns for safety observations (safe practices [SP], minor risk–no assist [MR], risk–potential harm [PH], and severe risk–prevent harm [SR]). The level of safety risk is circled for each task. Column 5 contains four columns to indicate quality of performance (standards met [SM], standards met–improvement possible [IP], standards partially met [PM], and standards not met [NM]). The level of quality is circled for each task. Column 6 contains three columns for the cumulative summary scores for task independence data, task safety data, and task quality data. Independence data is circled for each task, and a task summary score is entered for safety and quality based on the lowest subtask rating.

The summary scores described in **Table 2** were based on independence, safety, and quality data derived from columns 3–5 of the FEW-C data collection form. Summary scores for all item subtasks are rated on a predefined 4-point ordinal scale ranging from 0 to 3, with a score of 3 equal to independent, safe, and quality standards met (**Figure**). For independence data, the type and number of assists provided (VA, VSA, and PA—column 3, **Figure**)

*Schmeler MR. Development and testing of a clinical outcome measurement tool to assess wheeled mobility and seating interventions [dissertation]. Pittsburgh (PA): University of Pittsburgh; 2005.

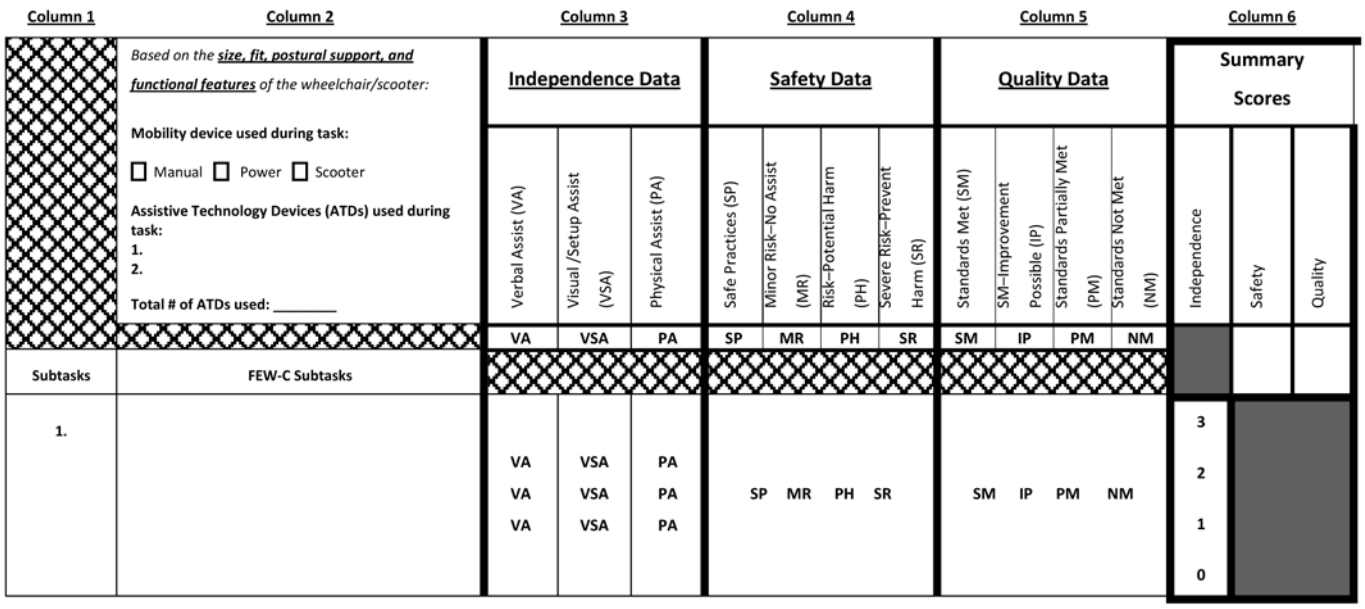

Figure.
Functioning Everyday with a Wheelchair-Capacity (FEW-C) example data collection form.

Table 2.
Functioning Everyday with a Wheelchair-Capacity scoring information.

Score	Independence Data	Safety Data	Quality Data
3	No assists given for task initiation, continuation, or completion.	**SP** = Safe practices were observed.	**SM** = Acceptable (standards met).
2	No physical assists given; 2 verbal or visual assists *or* 4 verbal or visual assists given.	**MR** = Minor risks were evident but no assistance provided.	**IP** = Acceptable (standards met, but improvement possible).
1	2 physical assists given but no total assistance; 3 verbal or visual assists; *or* 5 verbal or visual assists given.	**PH** = Risks to safety were observed and assistance given to prevent potential harm.	**PM** = Marginal (standards partially met).
0	3 physical assists given *or* total assistance required for task initiation, continuation, or completion.	**SR** = Severe risks evident: assistance provided to prevent harm.	**NM** = Unacceptable (standards not met).

was used to determine an independence summary score for each subtask. For example, for two VAs, one VSA, and one PA, the subtask summary score was 1. The type of safety risk observed (SP, MR, PH, and SR—column 4, **Figure**) for each subtask was used to derive a safety summary score for overall task performance. For example, the safety score for a subtask may have been MR (which corresponds to 2) or PH (which corresponds to 1). The highest level of safety risk observed or the lowest score (i.e., PH or 1) across all subtasks was used to determine the safety summary score so that consumers were not placed in a situation of risk through overestimation of task safety. For quality data, the standard of subtask performance (SM, IP, PM, and NM—column 5, **Figure**) was used to derive a quality summary score for overall task performance. For example, if the quality score for a subtask was PM (which corresponds to 1) and SM (which corresponds to 3) for another subtask, then the lowest quality level or the lowest score (i.e., PM or 1) across all subtasks was used to determine the quality summary score; in this way the consumer was not placed in a situation of risk through overestimation of task quality. The FEW-C has demonstrated good-to-excellent internal consistency, moderate to strong convergent and discriminant validity as well as excellent interrater reliability in a sample of adult manual and power wheelchair users.

IP-TR Reliability Study

The reliability of a tool speaks to its ability to yield consistent responses under given conditions [15]. For the current study, the more clinically relevant interrater reliability and internal consistency were examined. Interrater reliability was used to address the degree of consistency between IP and TR raters who observed and scored the FEW-C for each patient. Internal consistency was used to assess whether the component parts (items) of the FEW-C were measuring the same construct.

The four steps in achieving the IP-TR reliability study were—

1. Create training videos and conduct IP training sessions for administering and scoring the FEW-C.
2. Implement IP-TR interrater reliability testing using the FEW-C.
3. Establish ≥0.80 interrater reliability using the intraclass correlation coefficient (ICC) (two-way random average measures).
4. Establish the internal consistency of the FEW-C constructs of independence, safety, and quality.

Study Participant Procedures

The reliability study was conducted over a period of 26 months, during which time the remote clinics had their wheelchair clinic once every few months. The generalist practitioners, with the guidance of the research team, determined how the FEW-C was to be incorporated into their clinical service delivery model [16]. The FEW-C was administered to participants with their current WMS device after the initial intake. The generalist practitioners took the lead and asked the participants to perform each of the required FEW-C tasks. If there were questions or concerns about a specific task, the expert practitioner was there via TR to answer them. The expert practitioner never specifically told the generalist practitioners how to score a specific task as that was kept confidential. The expert practitioner via TR and the IP generalist practitioners on site observed the participants, and simultaneously rated each of the FEW-C tasks. For each patient, the FEW-C scores were compared between the expert practitioner and the on-site IP generalist practitioner. Each of the hospitals' wheelchair clinics was set up like an activities of daily living laboratory, with a sink, countertops, or a mirror for performing personal care tasks; linoleum flooring with a transition to carpet for indoor mobility and operation of WMS devices; and a mat table for transferring to and from WMS devices.

Study Raters

A total of five raters participated in the study: one expert practitioner and four generalist practitioners. The expert practitioner was an occupational therapist with an assistive technology professional certification and 10-plus years of clinical experience with WMS as well as one of the developers of the FEW-C. The four generalist practitioners were occupational therapists with varying levels of clinical experience and exposure to WMS.

Study Raters Training

Each of the four remote hospitals was at least 100 miles away from the University of Pittsburgh. Before the study started, the expert practitioner had a former client who used a power wheelchair with power seat functions come to the Center for Assistive Technology to be the subject in a training video on how to administer, score, and interpret the FEW-C outcome measure. The training video was sent to each of the generalist practitioners, along with the FEW-C test manual. The raters reviewed the training video and test manual before members of the research team traveled to each of the respective hospitals to conduct training sessions. A member of the research team simulated symptoms of an individual with multiple sclerosis and performed each of the FEW-C tasks. During the IP training sessions, the expert practitioner clarified any discrepancies and questions about the administration and scoring of the FEW-C. The generalist practitioners scoring of the FEW-C was consistent after their individualized training sessions. The training sessions lasted between 2 and 3 hours.

Apparatus

A set of videoconferencing equipment was sent to the information technology support staff at each of the remote clinics for setup. The system consisted of a Logitech QuickCam Orbit AF Web camera (Logitech; Fremont, California), a Panasonic network camera (model BB-HCM381A, Panasonic; Secaucus, New Jersey), and a CD with instructions on how to download and install Versatile and Integrated System for Telerehabilitation (VISYTER), a custom, secure, Internet Protocol-based videoconferencing system developed within the Rehabilitation Engineering Research Center (RERC) on TR. The Logitech QuickCam Orbit AF Web camera was placed on the desk next to the personal computer and used to initiate the video and audio communication within VISYTER. The Panasonic network camera was mounted within the laboratory space and integrated within VISYTER, allowing the expert

practitioner another camera angle and the ability to remotely control the pan, tilt, and zoom settings. The remote clinics were connected to the Internet using either broadband (ADSL, 728 Kbps up/1.5 Mbps down) or T1 (1.544 Mbps) connections and using standard quality video (320 × 240, 30 frames per second). The lighting intensity at the sites was between 750 and 1,000 lux, while the walls were light gray or a pale blue with a flat finish to assist with contrast. Researchers worked closely with the information technology staff from the remote clinics to deploy VISYTER. The most common task was to ensure that the settings from both networks (i.e., metropolitan area clinic and remote clinic) allowed a direct connection between the participants for a real-time teleconsultation session. In this task, the network requirement functions, such as opening ports in the firewall, creating forwarding protocols to allow both ends of the communication line to recognize each other, and filtering packages to allow only connections from known sources, were reviewed and implemented. A set of trials was conducted between the University of Pittsburgh and each one of the remote hospitals to configure these functions along with both the audio and video settings before data collection was initiated. The generalist practitioners also received training on how to set up the equipment in case the information technologists were not available.

Data Analysis

The data were analyzed using SPSS, version 14.0 (SPSS Inc; Chicago, Illinois). Descriptive statistics were used to determine the frequencies for all the variables recorded in the data collection, including age, sex, primary diagnosis, race/ethnicity, and current WMS device.

According to Portney and Watkins, interrater reliability is best assessed when raters are able to measure a response during a single trial in which they can observe a subject simultaneously and independently [15]. Test-retest reliability was not performed because of study participant and family member availability and burden. We established interrater reliability using the ICC by selecting the two-way effects model and absolute agreement menus within SPSS with a target of ≥0.80 [15,17]. The internal consistency of the total FEW-C tool was examined using Cronbach alpha. Cronbach alpha reflects the extent to which item responses correlate with each other and with a total test score. An alpha >0.70 but <0.95 was set as the statistically acceptable standard coefficient because it would indicate good to excellent homogeneity of the total FEW-C, without unnecessary redundancy of items [15].

RESULTS

Interrater Reliability

The IP-TR data demonstrated excellent interrater reliability with an ICC of 0.91 (95% confidence interval [CI]: 0.84–0.95, $p < 0.001$). This measure was also consistent for each of the items, in which the combined ICC data and summary scores were ≥0.80 for independence, safety, and quality (**Table 3**). Both of these primary findings were above the acceptable target value of ≥0.80, and all reliability coefficients had small to moderate CIs, indicating that TR had good precision in determining capacity-based testing with an IP generalist practitioner administering the FEW-C. With all ICCs ≥0.80, the ratings also

Table 3.
Interrater reliability of the Functioning Everyday with a Wheelchair-Capacity (FEW-C) via telerehabilitation. Data are presented as intraclass correlation coefficient (95% confidence interval).

FEW-C Task	Overall	I	S	Q
Comfort	0.91 (0.83–0.95)	0.99 (0.98–0.99)	0.88 (0.78–0.93)	0.86 (0.74–0.92)
Health Needs	0.89 (0.79–0.94)	0.96 (0.92–0.98)	0.85 (0.72–0.92)	0.85 (0.72–0.92)
Operate	0.91 (0.85–0.95)	0.93 (0.88–0.96)	0.91 (0.84–0.95)	0.90 (0.82–0.95)
Reach	0.88 (0.79–0.94)	0.93 (0.87–0.96)	0.85 (0.73–0.92)	0.87 (0.76–0.93)
Transfers	0.91 (0.82–0.95)	0.98 (0.96–0.99)	0.83 (0.68–0.90)	0.91 (0.83–0.95)
Personal Care	0.95 (0.91–0.97)	0.96 (0.93–0.98)	0.93 (0.87–0.96)	0.96 (0.93–0.98)
Indoor Mobility	0.92 (0.87–0.96)	0.98 (0.97–0.99)	0.89 (0.80–0.94)	0.90 (0.83–0.95)
Total	0.91 (0.84–0.95)	0.96 (0.93–0.98)	0.88 (0.77–0.93)	0.90 (0.82–0.94)

I = independence, Q = quality, S = safety.

indicated the reliability of the video camera and video-conferencing system quality for allowing the expert practitioner to rate the FEW-C tasks via TR at the same time that the generalist practitioner was seeing and rating the participant IP.

Internal Consistency

Internal consistency of the total FEW-C tool achieved a standardized alpha of 0.94. Internal consistency for each scale was also good, with standardized alphas of 0.91, 0.83, and 0.82 for independence, safety, and quality, respectively (**Tables 4–6**). Internal consistency data was derived from the scores of the expert practitioner who observed and scored participants via TR.

DISCUSSION

Using the FEW-C performance-based outcome tool, we were able to demonstrate excellent IP-TR interrater reliability as well as internal consistency. Tools for assessing clinical performance and health outcomes have progressed considerably in recent years as methodologists and researchers have tested and improved the reliability and validity of measures and made them more relevant and usable in routine clinical practice. Nevertheless, determination of the interrater reliability between an IP generalist practitioner and an expert practitioner using TR has been limited in peer-reviewed literature [11–14]. TR provided a means of evaluating the participants from more than 100 miles away as they performed specific tasks in their WMS devices. The lack of "hands on" exposure is just one barrier associated with TR. Nevertheless, our data confirmed that an expert practitioner can assess an individual's functional status remotely and with as much accuracy as a general practitioner who assesses the same individual IP.

These results are comparable to the results found during the validation of the FEW-C, in which the raters observed 15 IP FEW-C administrations compared with the 46 in this study.* Schmeler reported that the FEW-C demonstrated excellent interrater reliability with an ICC of 0.99 (95% CI: 0.98–0.99, $p < 0.001$), whereas this study

*Schmeler MR. Development and testing of a clinical outcome measurement tool to assess wheeled mobility and seating interventions [dissertation]. Pittsburgh (PA): University of Pittsburgh; 2005.

Table 4.
Independence construct: Internal consistency of Functioning Everyday with a Wheelchair-Capacity (FEW-C) via telerehabilitation.

FEW-C Task	COM	HN	OP	RCH	TRN	PC	IM
COM	1.00	0.78	0.66	0.61	0.50	0.44	0.56
HN	—	1.00	0.55	0.69	0.65	0.59	0.45
OP	—	—	1.00	0.59	0.46	0.57	0.87
RCH	—	—	—	1.00	0.64	0.61	0.47
TRN	—	—	—	—	1.00	0.67	0.41
PC	—	—	—	—	—	1.00	0.56
IM	—	—	—	—	—	—	1.00
Overall	—	—	—	—	—	—	0.91

COM = comfort, HN = health needs, IM = indoor mobility, OP = operate, PC = personal care, RCH = reach, TRN = transfers.

Table 5.
Safety construct: Internal consistency of Functioning Everyday with a Wheelchair-Capacity (FEW-C) via telerehabilitation.

FEW-C Task	COM	HN	OP	RCH	TRN	PC	IM
COM	1.00	0.70	0.22	0.35	0.26	0.29	0.35
HN	—	1.00	0.31	0.33	0.50	0.40	0.31
OP	—	—	1.00	0.47	0.32	0.39	0.67
RCH	—	—	—	1.00	0.51	0.58	0.52
TRN	—	—	—	—	1.00	0.47	0.34
PC	—	—	—	—	—	1.00	0.47
IM	—	—	—	—	—	—	1.00
Overall	—	—	—	—	—	—	0.83

COM = comfort, HN = health needs, IM = indoor mobility, OP = operate, PC = personal care, RCH = reach, TRN = transfers.

Table 6.
Quality construct: Internal consistency of Functioning Everyday with a Wheelchair-Capacity (FEW-C) via telerehabilitation.

FEW-C Task	COM	HN	OP	RCH	TRN	PC	IM
COM	1.00	0.72	0.30	0.34	0.37	0.22	0.44
HN	—	1.00	0.24	0.35	0.28	0.35	0.25
OP	—	—	1.00	0.45	0.40	0.22	0.62
RCH	—	—	—	1.00	0.55	0.58	0.46
TRN	—	—	—	—	1.00	0.41	0.33
PC	—	—	—	—	—	1.00	0.42
IM	—	—	—	—	—	—	1.00
Overall	—	—	—	—	—	—	0.82

COM = comfort, HN = health needs, IM = indoor mobility, OP = operate, PC = personal care, RCH = reach, TRN = transfers.

resulted in an ICC of 0.91 (95% CI: 0.84–0.95, $p < 0.001$). In looking at each item of independence, safety, and quality, both studies reached the acceptable value of ≥0.80. The results were also consistent for internal consistency, with both studies achieving a standardized alpha of >0.70 but <0.95 for independence, safety, and quality.

Until recently, equipment costs were prohibitively high and bandwidth issues hampered communication. Now with improvements in Web cameras and broadband high-speed Internet access, these problems have been solved. After the study, the expert practitioner noted how important live interaction was and that merely videotaping a patient and sending the video may not provide enough information to make a recommendation. The VISYTER videoconferencing system enabled the expert practitioner to request a closer view of specific FEW-C tasks or ask follow-up questions when needed. Furthermore, the expert practitioner felt that the remote assessment was the same as what a patient would receive in Pittsburgh. The only difference was that the generalist practitioner took care of the hands-on portions of the assessment and was the treating clinician on record. In regard to the videoconferencing system, the expert practitioner stated that the quality of both the video and audio being sent from the remote sites was acceptable and of a high quality and the generalist practitioners at the remote sites agreed.

Several factors contributed to the excellent interrater reliability between our raters. First, there are two classes of TR techniques: synchronous (i.e., real-time) and asynchronous (i.e., store-and-forward or non-real-time) interactions. The ability of the expert practitioner to simultaneously (in real time) observe and score the participant via TR while the generalist practitioner was seeing the participant IP eliminated the problems of patient fatigue and allowed the generalist practitioner to clarify procedures with the expert practitioner if necessary. Second, the systematic training procedures implemented before the study started provided the generalist practitioners at the remote sites with training videos and the FEW-C training manual to prepare for the IP training sessions. Also, members of the research team and all raters participated in on-site training sessions conducted by one of the primary developers of the FEW-C instrument. A third contributing factor was the generalist practitioners and information technologists at each of the remote sites. The generalist practitioners were all licensed occupational therapists with varying levels of clinical experience and limited experience in the area of WMS. These practitioners were eager to learn how to implement the FEW-C as part of the WMS assessment process. Similarly, the information technologists at each site were excellent and verified and tested both the video and audio quality of the videoconferencing system before each scheduled clinic visit.

One primary limitation of the study was the number of raters. There were only two raters during each administration of the FEW-C. Future investigation of the FEW-C should take advantage of the asynchronous or store-and-forward interaction capabilities of the videoconferencing system. The capability to record and archive each of the administrations allows for future studies to be conducted for either educational or research purposes. Furthermore, the videos can be used for training purposes and establishing protocols; then videos can be posted to a Web site where practitioners can watch and score how an individual performs the FEW-C tasks and interrater reliability can be conducted with a larger set of raters. Although test-retest reliability would be desirable, similar to other reliability studies [14], we did not include it because of the burden it would place on participants, who would have to return to the clinics for additional visits.

Obtaining quantitative data with thorough methodologies is important not only for demonstrating the value of TR but also for assisting with policy formulation. The demonstration of increased provider efficiency and patient outcomes is needed to address policy issues such as reimbursement and licensure as well as to support increased use of TR as a service delivery model when appropriate. When this research study was initiated, only a few resources defined TR. Currently, with additional peer-reviewed publications, conference proceedings, position papers, and clinical guidelines from professional associations and organizations, the need for and utility of TR has been recognized amongst its peers.

CONCLUSIONS

Using TR, an expert practitioner had the ability to observe and rate patient function with the same degree of accuracy as the IP generalist practitioner at a remote site more than 100 miles away. These findings indicate that TR is a valuable adjunct and support in the specialized area of WMS assessment, especially when a valid and reliable performance-based measure is used. Choosing an observation-based outcome measure that properly meets the participants' needs, settings, services, and desired outcomes is important for a TR assessment. In any service delivery model, psychometrically sound outcome measures are a necessity. The FEW-C met the psychometric

criteria and was implemented into the standards of practice for performing TR WMS assessments at four remote sites. Future development of valid and reliable assessment tools appropriate for TR could improve the quality of WMS and other rehabilitation services, as well as develop the skills and confidence of generalist practitioners in remote rehabilitation clinics.

ACKNOWLEDGMENTS

Author Contributions:
Study concept and design: R. M. Schein, M. R. Schmeler, M. B. Holm, D. M. Brienza.
Acquisition of data: R. M. Schein, M. R. Schmeler.
Analysis and interpretation of data: R. M. Schein, M. B. Holm, M. Pramuka.
Drafting of manuscript: R. M. Schein, M. B. Holm, M. Pramuka, D. M. Brienza.
Statistical analysis: R. M. Schein, M. B. Holm.
Obtained funding: M. R. Schmeler, D. M. Brienza.
Administrative, technical, or material support: A. Saptono.
Study supervision: M. R. Schmeler.
Financial Disclosures: The authors have declared that no competing interests exist.
Funding/Support: This material was based on work supported by the National Institute on Disability and Rehabilitation Research, United States Department of Education (grant H133E040012), at the University of Pittsburgh RERC on TR (www.rerctr.pitt.edu).
Institutional Review: The institutional review boards of each of the four remote hospitals (DRMC, ERHC, CCMH, MMCHS) and the University of Pittsburgh reviewed and approved the study protocols.
Participant Follow-Up: The authors do not plan to notify participants of the publication of this article because of a lack of contact information.

REFERENCES

1. Bashshur RL. Telemedicine/telehealth: An international perspective. Telemedicine and health care. Telemed J E Health. 2002;8(1):5–12. [PMID: 12020402] DOI:10.1089/15305620252933365
2. Lemaire ED, Boudrias Y, Greene G. Low-bandwidth, Internet-based videoconferencing for physical rehabilitation consultations. J Telemed Telecare. 2001;7(2):82–89. [PMID: 11331045] DOI:10.1258/1357633011936200
3. Torsney K. Advantages and disadvantages of telerehabilitation for persons with neurological disabilities. NeuroRehabilitation 2003;18(2):183–85. [PMID: 12867681]
4. Winters JM. Telerehabilitation research: Emerging opportunities. Annu Rev Biomed Eng. 2002;4:287–320. [PMID: 12117760] DOI:10.1146/annurev.bioeng.4.112801.121923
5. Hoenig HJ, Sanford JA, Butterfield T, Griffiths PC, Richardson P, Hargraves K. Development of a teletechnology protocol for in-home rehabilitation. J Rehabil Res Dev. 2006; 43(2):287–98. [PMID: 16847794] DOI:10.1682/JRRD.2004.07.0089
6. Rosen MJ. Telerehabilitation. Telemed J E Health. 2004; 10(2):115–17. [PMID: 15319039] DOI:10.1089/tmj.2004.10.115
7. Schmeler MR, Schein RM, McCue MP, Betz K. Telerehabilitation and clinical applications: Research, opportunities, and challenges. Int J Telerehabil. 2009;1:59–72.
8. Malagodi M, Schmeler MR, Shapcott NG, Pelleschi T. The use of telemedicine in assistive technology service delivery: Results of a pilot study. Tech Spec Interest Section Q. 1998;8.
9. Lemaire ED, Jeffreys Y, Morazain G, Greene G. A model for remote orthotic assessment using distance communication technology. Saudi J Dis Rehabil. 1997;3:1–14.
10. Cooper R, Fitzgerald S, Boninger M, Cooper RA, Shapcott N, Cohen L, Thorman T, Schmeler MR. Using telerehabilitation to aid in selecting a wheelchair. Proceedings of the RESNA 2002 Annual Conference; 2002 Jun 27–Jul 1; Minneapolis, MN. Washington (DC): RESNA Press; 2002. p. 327.
11. Allegretti A, Fitzgerald S, Schmeler MR, Cooper RA, Boninger M, Shapcott N. Pelvic positioning evaluations for wheelchair selection: A comparison between in-person and video conferencing. Proceedings of the 20th International Seating Symposium; 2004 Mar 4–6; Vancouver, Canada. p. 215–16.
12. Dreyer NC, Dreyer KA, Shaw DK, Wittman PP. Efficacy of telemedicine in occupational therapy: A pilot study. J Allied Health. 2001;30(1):39–42. [PMID: 11265272]
13. Shafqat S, Kvedar JC, Guanci MM, Chang Y, Schwamm LH. Role for telemedicine in acute stroke. Feasibility and reliability of remote administration of the NIH stroke scale. Stroke. 1999;30(10):2141–45. [PMID: 10512919]
14. Palsbo SE, Dawson SJ, Savard L, Goldstein M, Heuser A. Televideo assessment using Functional Reach Test and European Stroke Scale. J Rehabil Res Dev. 2007;44(5): 659–64. [PMID: 17943677] DOI:10.1682/JRRD.2006.11.0144
15. Portney LG, Watkins MP. Foundations of clinical research: Applications to practice. Upper Saddle River (NJ): Prentice Hall; 2000.
16. Schein RM, Schmeler MR, Brienza D, Saptono A, Parmanto B. Development of a service delivery protocol used for remote wheelchair consultation via telerehabilitation. Telemed J E Health. 2008;14(9):932–38. [PMID: 19035803] DOI:10.1089/tmj.2008.0010
17. Huck SW, Cormier WH. Reading statistics and research. 2nd ed. New York (NY): Harper Collins; 1996.

Submitted for publication March 17, 2010. Accepted in revised form October 12, 2010.

This article and any supplementary material should be cited as follows:
Schein RM, Schmeler MR, Holm MB, Pramuka M, Saptono A, Brienza DM. Telerehabilitation assessment using the Functioning Everyday with a Wheelchair-Capacity instrument. J Rehabil Res Dev. 2011;48(2):115–24. DOI:10.1682/JRRD.2010.03.0039