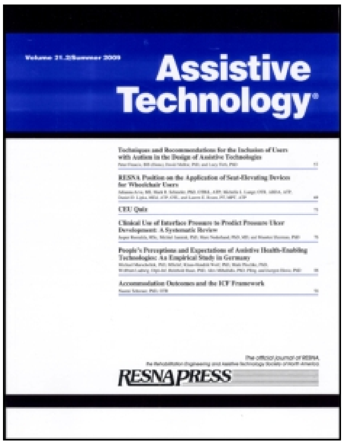


## Assistive Technology

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t910220176

# Validation of the Seating and Mobility Script Concordance Test

Laura J. Cohen[a]; Shirley G. Fitzgerald[b]; Suzanne Lane[c]; Michael L. Boninger[d]; Jean Minkel[e]; Michael McCue[f]

[a] Shepherd Center, Crawford Research Institute, Atlanta, Georgia, USA [b] James A. Haley Tampa VA, Tampa, Florida, USA [c] School of Education, Department of Psychology in Education, University of Pittsburgh, Pittsburgh, Pennsylvania, USA [d] School of Medicine, Department of Physical Medicine & Rehabilitation, University of Pittsburgh, and Human Engineering Research Laboratories, VA Pittsburgh Healthcare System, Pittsburgh, Pennsylvania, USA [e] Minkel Consulting, New Windsor, New York, USA [f] School of Health and Rehabilitation Sciences, Department of Rehabilitation Science and Technology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

**To cite this Article** Cohen, Laura J. , Fitzgerald, Shirley G. , Lane, Suzanne , Boninger, Michael L. , Minkel, Jean and McCue, Michael(2009) 'Validation of the Seating and Mobility Script Concordance Test', Assistive Technology, 21: 1, 47 — 56

**To link to this Article: DOI:** 10.1080/10400430902945546

**URL:** http://dx.doi.org/10.1080/10400430902945546

***Assistive Technology***, 21:47–56, 2009

ISSN: 1040-0435 print
DOI: 10.1080/10400430902945546





# Validation of the Seating and Mobility Script Concordance Test

**Laura J. Cohen, PT, PhD, ATP,[1]**
**Shirley G. Fitzgerald, PhD,[2]**
**Suzanne Lane, PhD,[3]**
**Michael L. Boninger, MD,[4]**
**Jean Minkel, MA, PT,[5] and**
**Michael McCue, PhD[6]**
[1]Crawford Research Institute, Shepherd Center, Atlanta, Georgia
[2]James A. Haley Tampa VA, Tampa, Florida
[3]School of Education, Department of Psychology in Education, University of Pittsburgh, Pittsburgh, Pennsylvania
[4]School of Medicine, Department of Physical Medicine & Rehabilitation, University of Pittsburgh, and Human Engineering Research Laboratories, VA Pittsburgh Healthcare System, Pittsburgh, Pennsylvania
[5]Minkel Consulting, New Windsor, New York
[6]School of Health and Rehabilitation Sciences, Department of Rehabilitation Science and Technology, University of Pittsburgh, Pittsburgh, Pennsylvania

**ABSTRACT** The purpose of this study was to develop the scoring system for the Seating and Mobility Script Concordance Test (SMSCT), obtain and appraise internal and external structure evidence, and assess the validity of the SMSCT. The SMSCT purpose is to provide a method for testing knowledge of seating and mobility prescription. A sample of 106 therapists and 15 spinal cord injury experts contributed to the development of the scoring system. Validity evidence was obtained using 15 seating and mobility experts, 10 orthopedic experts, and 66 therapists with varying levels of seating and mobility expertise. Proxy measures of clinical expertise were used for external validity evidence since no criterion measures exist. The SMSCT was found to differentiate between seating and mobility experts' and orthopedic experts' intervention subtest scores ($p = 0.04$). The proxy measure of clinical expertise, seating and mobility hours/week, was found to predict SMSCT intervention scores ($p = 0.002$). The internal structure of the SMSCT may include evidence of reduced item performance but satisfactory convergent and discriminate evidence by construct definition. Although the SMSCT may be a promising approach for measuring seating and mobility expertise, limitations exist in the corrected content. Future application of the SMSCT should only be used after further development of the tool occurs.

**KEYWORDS** clinical competence, educational measurement, professional practice, rehabilitation, validation

## INTRODUCTION

The appropriateness of a consumer's seating and mobility (SM) system varies considerably depending on the competence, proficiency, and experience of the professionals assisting the user (Herman & Lange, 1999). At present, there is a scarcity of skilled and knowledgeable therapists to evaluate and recommend SM devices (Herman & Lange, 1999; Fifield & Fifield, 1997; Beaulieu, 2007). With ever-changing and emerging technologies available in the area of SM, it is necessary that clinicians continually update their knowledge, skills, and clinical competencies in order to provide quality care. The most widely accepted way to upgrade professional training is participation in continuing education activities. Yet, the effectiveness of professional training is still

Address correspondence to Shirley G. Fitzgerald, PhD, HSR&D REAP, 8900 Grand Oak Circle (118 M), Tampa, FL 33637.
E-mail: Shirley.Fitzgerald@va.gov

unknown. While the need to train more skilled practitioners is clear, the tools to assess the effectiveness of training programs and the most effective means of training have yet to be identified (Davis, Thomson, Oxman, & Haynes, 1992; Grossman, 1998).

Organization of knowledge (Chi & Glaser, 1988; Charlin, Roy, Brailovsky, & van der Vleuten, 2000; Schmidt, Norman, & Boshuizen, 1990) is a key characteristic in experts' performance that has been found to be robust and generalizable across various domains of research, including areas of cognitive psychology and cognitive sciences, and it is viewed as a clinical determinant of expertise in medicine. Understanding the way experts practice, think, solve problems, and operate is necessary in order to define essential skills and clinical competencies. This is also necessary in promoting the continued preparation of practicing professionals and the development of the next generation of professionals (Jensen, Gwyer, Shepard, & Hack, 2000).

Over the last 30 years, there has been an active line of research and an accumulation of evidence in medicine about the process of how knowledge is (a) initially learned, (b) organized in memory, (c) accessed later to solve problems, and (d) modified with experience (Charlin et al., 2000; Schmidt et al., 1990). Authors have hypothesized that in clinical medicine, differences between experts and novices lay primarily in experts' recall of meaningful relationships and patterns, that is, the structure of knowledge rather than the problem-solving strategy applied to the problem (Charlin et al., 2000; Schmidt et al., 1990). The acquisition of expertise in a content area can be characterized by the development of distinctive memory structures called *scripts*, which are meaningful sets of connections among abstract concepts and/or specific experiences (Charlin et al., 2000; Schmidt et al., 1990; Tardif & Boshuizen, 2000). Information such as the assumptions and hypotheses that are necessary to assess and manage clinical cases is retrieved through the activation of these scripts. Research in this area is attempting to portray how a script as a memory structure might be organized for specific clinical tasks. Thus, it follows that in testing clinical expertise, it is necessary to access these scripts by using specific, relevant clinical situations in the test items in order to assess the organization of knowledge.

At present, no validated measurement tests exist to evaluate the impact of educational experiences or clinical practice on the ability to make specialized clinical decisions about SM needs. The Seating and Mobility Script Concordance Test (SMSCT) is a tool designed to assess clinicians by examining the organization of their knowledge, associations between items addressing their knowledge, and the adequacy of their clinical decisions compared to expert consensus. The SMSCT is modeled after the development of a similar script concordance test (SCT) in the field of medicine created by Charlin et al. (2000). The SMSCT consists of 67 items divided into two subtests: 33 assessment items (49%) and 34 intervention items (51%) (Cohen, Fitzgerald, Lane, & Boninger, 2005). As knowledge in the area of seating and mobility can vary dependent on the disability, it was decided to focus the subject matter for the SMSCT on SM for individuals with spinal cord injuries (SCIs).

Several phases were dedicated to the development and validation of the SMSCT (Cohen et al., 2005; Cohen, Fitzgerald, Trefler, Boninger, & McCue, 2002). To date, only the content validity evidence of the SMSCT has been investigated (Cohen et al., 2005). Determining the validity of the SMSCT is the most fundamental consideration in evaluating its usefulness as a measure of clinical expertise for clinicians who work with SCIs. The purpose of this study was twofold: first, to provide evidence to support the internal structure of the SMSCT items, and, second, to obtain external structure evidence to support how the SMSCT is interpreted. We had two specific aims: (a) the development of the scoring system and (b) the collection and appraisal of internal and external structure evidence. We sought evidence to explore the following research hypotheses (the first linked to our first aim and the others linked to our second aim):

*Hypothesis 1:* There is a significant relationship between each item response and the total SMSCT score.

*Hypothesis 2:* There is a significant relationship among item responses within a subtest as compared to item responses in other subtests (assessment, intervention).

*Hypothesis 3:* SMSCT subscores can differentiate between SM experts (the SM group) and orthopedic experts (the Ortho group).

*Hypothesis 4:* Proxy measures of clinical expertise (i.e., years of clinical practice, years of seating and mobility provision, hours per week of seating and mobility services, number of spinal cord–injured patients treated per year, professional-level degree) will predict SMSCT subscores.

# METHODS

## Participants

Overall, 115 clinicians agreed to participate in this project, of whom 15 contributed to the development of the SMSCT scoring system. The remaining 100 were recruited for acquiring validity evidence. Figure 1 illustrates the subject groups used for this work. All subjects signed a consent form approved by the internal review board.

For the purpose of developing the SMSCT scoring system, 15 expert SM clinicians were recruited from different model SCI systems across the United States. These clinicians, hereafter referred to as SCI experts, were selected because they regularly recommend SM equipment to individuals with SCIs. The number of experts was based on the work of Charlin et al. (2000), which stated that the number of experts used to develop a similar SCT scoring system must be sufficient to express the variability in answers that experts may show for each item. Their work suggests a sample size of 9–12 experts (Charlin et al., 2000; Charlin, Desaulniers, Gagnon, Blouin, & van der Vleuten, 2002). The model SCI systems were used for expert recruitment, since therapists from these settings would most likely meet the eligibility criteria specific to the content area of SCIs. It is recognized that this sample may not be representative of the SM population. For the purposes of this work, we defined expert SM clinicians as individuals with:

- A physical or occupational therapy license
- A combination of SM service provision equating to full-time work (approximately 40 hours per week) for 5 years
- Completion of professional development (i.e., continuing education courses, manufacturer in-services, graduate course work, etc.) to include a minimum of 10 contact hours per year for a minimum of 5 years in the area of SM as documented by self-report

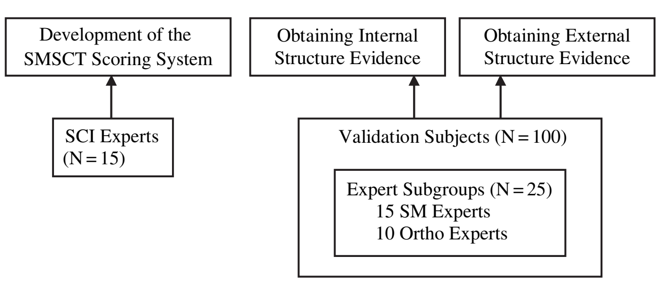


**FIGURE 1** Subject groups and study components.

In order to obtain external validity evidence and compare groups of known qualities, a subset of the 100 validation subjects, 15 experts in SM and 10 orthopedic physical therapy (PT) experts, were recruited separately. The SM experts were selected because they differed from the SCI experts who developed the scoring system in that they worked with a diverse patient population not exclusive to SCIs. The orthopedic experts were chosen in order to recruit a homogeneous group that had expertise related to the musculoskeletal spine but not specific to SM service provision.

The SCI and SM experts were recruited through invitation by the investigators, whereas the orthopedic experts were recruited through invitation by the president and delegate of the Research Section of the American Physical Therapy Association. Volunteer PT and occupational therapy (OT) clinicians with differing levels of SM experience made up the remainder of the validation group. All validation subjects were licensed therapists. These subjects were recruited in person at the International Seating Symposium, the Rehabilitation Engineering and Assistive Technology Society of North America (RESNA) annual conference, and, to a limited extent, by word of mouth.

The correlation coefficient for the test is influenced by the total number and the diversity in experience of the respondents. Anticipating a conservative minimal effect correlation of $r = .3$ and $\alpha = .05$, 100 subjects were required to achieve a power of 0.87 (Cohen, 1988), thus reducing the likelihood of a Type II error.

## SMSCT Test Administration Procedures and Data Management

All subjects (SCI experts and validation subjects) completed a demographic questionnaire and the 67-item SMSCT. Demographic information included gender, profession, professional-level degree, years of clinical practice, years of SM service provision, number of SCI patients seen per year, type of AT training, diagnoses commonly treated, and professional development activities. Professional-level degree was defined as the clinicians' highest degree acquired to practice physical or occupational therapy (i.e., entry-level associates, bachelor's, master's, or clinical doctorate versus advanced-level master's, clinical doctorate, or PhD).

Subjects took the SMSCT in a test room setup, without the ability to consult with their peers. The subjects were required to record their answers on a machine-scored answer sheet. The test took approximately 60 minutes to complete. All demographic data were entered into a Microsoft Access database, coded, and verified for accuracy. SMSCT data were processed by the Office of Measurement and Testing at the University of Pittsburgh, resulting in an electronic document that was used for scoring and further analysis by the research team.

## Procedures for the Development of the Scoring System

The scoring system for the SMSCT is modeled after Charlin et al.'s (2000) innovative work. The scoring system for the SMSCT is founded on the principle that any answer given by an expert has an intrinsic value, even if other experts do not agree with it. Hence, scores for each item of the SMSCT are computed from the frequencies given to each point of the Likert scale by the experts (Charlin et al., 2000). We chose to recruit a total of 15 SCI experts for the development of the scoring system in order to allow for attrition. Previous research has established that experts' answers vary when they have to solve clinical problems, even in their own field of expertise (Charlin et al., 2000; Norman, 2000). Other studies using the test format for an SCT support this finding and show that experts provide the same answers on some items but provide different answers on others (Charlin et al., 2000; Beausoleil & van der Vleuten, 2001; Charlin et al., 1998).

The item responses of the 15 SCI experts were used to prepare the scoring key based on the preestablished process (Charlin et al., 2000). For each item, answers were assigned a weighted value corresponding to the proportion of experts who selected the response. An example of the item format is shown in Figures 2 and 3; item scoring is illustrated in Table 1. The score for each question is the proportion of experts who gave the same answer for the question and is weighted by the degree of agreement between experts. The modal answer, or most frequently selected answer chosen by the experts, was used to transform item raw scores so that the modal expert response on each item received a maximum score of 1 and other experts' choices received a partial credit score. Answers not chosen by experts received a score of 0. The total score for a specific test is the sum of credit obtained on each item for each

For assessment knowledge

A 32-year-old male with a diagnosis of T10 spinal cord injury arrives in your clinic sitting in his manual wheelchair in a slumped kyphotic posture with a left ischial pressure ulcer.

| | If you were thinking of (a hypothesis) | And then you find | This hypothesis |
|---|---|---|---|
| 1 | Fixed posterior pelvic tilt | Supine mat assessment reveals increased lumbar lordosis | 1 2 3 4 5 |
| 2 | Fixed pelvic obliquity with right ASIS (anterior superior iliac spine) higher than the left | Supine mat assessment reveals full lower trunk mobility with left lateral flexion to create right lateral trunk extension | 1 2 3 4 5 |
| 3 | Fixed pelvic obliquity with right ASIS higher than the left | Left hip flexion range of motion 10 from neutral to 95 degrees | 1 2 3 4 5 |

1 becomes almost eliminated
2 becomes less probable
3 is not affected by the new information
4 becomes more probable
5 becomes most likely probable

**FIGURE 2** SMSCT item format: example assessment items.

For intervention knowledge

A 45- year-old man 20 years status post-C6-7 spinal cord injury is referred to your clinic for a replacement wheelchair. He is an experienced manual wheelchair user and is currently using a 10-year-old, folding frame, depot manual wheelchair with a fixed seat to back angle and no rear wheel axle adjustability. This chair presents with vinyl upholstery that is overstretched and results in his body being positioned between the back posts. He has removed the armrests for easier wheel access. He sits with a slumped posture with rounded shoulders and forward head position. His main complaint is neck and shoulder pain of recent slow onset (less than 4 months). He reports an active lifestyle including driving a car, independently loading/unloading his folding wheelchair, and employment as an architect. He lives in an accessible home environment with his wife and two children 5 and 7 years old. He states he is interested in trying new things that may alleviate his current problems or improve his pain.

| | **If you find** | **And then the supplier recommends the following** | **To what degree of confidence could you justify this recommendation?** |
|---|---|---|---|
| 1 | He reports that his neck and shoulder pain is less severe when propelling his wheelchair | Duplicating features of current wheelchair with a lightweight wheelchair | 1 2 3 4 5 |
| 2 | When sitting on the mat he falls forward and he puts his hands on the mat in order to stop himself | Adjustable tension back upholstery | 1 2 3 4 5 |
| 3 | Complaints of pain with shoulder extension, abduction, and internal rotation | Ultralight manual wheelchair with rear axle adjustability positioned in a mid to forward position | 1 2 3 4 5 |

1 with very little confidence
2 with little confidence
3 neither favor nor oppose the recommendation
4 with partial confidence
5 with a high degree of confidence

**FIGURE 3** **SMSCT item format: example intervention items.**

**TABLE 1** **Examples of item scoring**

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Experts' answers (No.) | 0 | 0 | 0 | 7 | 2 |
| Raw score | 0 | 0 | 0 | 7/9 (.77) | 2/9 (.22) |
| Modal transformed score | 0 | 0 | 0 | 7/7 | 2/7 |
| Credit for the item | 0 | 0 | 0 | 1 | .29 |

subject. Finally, the total score is transformed to obtain a maximum test score of 100 for ease of interpretation. A score of 100 signifies that the subject provides on each item the answer that most SCI experts provide, and the lower the score the farther the examinees are from the SCI experts' prototypic script for the situation.

## Procedures for Obtaining Internal Structure Evidence

To determine the level of internal consistency, item analyses were explored using the responses of the SCI experts ($n = 15$). Analyses included the study of the intercorrelation matrix, reliability coefficients, and corrected item/total correlations. We first examined the intercorrelation matrix to identify poorly performing items. Next we studied the alpha coefficient estimates used to measure item internal consistency. Internal consistency is affected by homogeneity of items: The more homogeneous the items, the higher the test score reliability. A stepwise reliability analysis was completed to exclude, one by one, all of the variables having a negative impact on reliability so that the best-performing items could be selected. Lastly, to determine if the two domains (assessment and intervention) were indeed measuring different constructs, we explored correlations for items within a subtest and items across subtests. The results from these analyses, combined with the judgments of the research team, guided decisions about which items were retained and/or eliminated from the final version of the SMSCT.

## Procedures for Obtaining External Structure Evidence

Independent-samples *t* tests were used to determine if the SMSCT subscores could differentiate between the SM group ($n = 15$) and the Ortho group ($n = 10$). To explore which proxy measures of clinical expertise could predict SMSCT subscale scores (assessment and intervention), bivariate correlations were examined. A subsequent stepwise linear multiple regression analysis was conducted to identify which factors could predict SMSCT subtest scores (assessment, intervention). The predictor variables (proxy measures of clinical expertise) were professional level, number of SCI patients per year, years of clinical practice, years of SM provision, and hours per week of SM. The criterion variables were SMSCT subtest scores (assessment or intervention). Two models were explored, first using the assessment scores as the criterion variable and next using the intervention scores as the criterion variable.

## Methods of Analysis Used for Obtaining Validity Evidence

Initial analyses of data were completed to look at data normalcy, outliers, and summary statistics. Descriptive statistics were used to describe group demographics. Frequency counts were used with nominal data, and measures of central tendency were used for continuous data. For all analyses, parametric statistics were used for normally distributed variables, homogeneity of group variances were estimated using the Levene test, nonparametric statistics were used to analyze data not normally distributed, and significance levels were set at $p < .05$. For stepwise multiple regression analyses, a stepping method criterion for entry was set at .05, and removal was set at .10.

# RESULTS

## Sample Demographics

Table 2 provides an overview of the subjects who participated in all phases of the study. As can be seen, the populations for each component primarily consisted of middle-aged (age range: 36–48 years) females. It should be noted that 8 subjects of 100 who enrolled in the study and planned to complete the test remotely withdrew from the study due to time constraints; one subject opted to withdraw due to lack of experience in SM.

## Internal Structure Evidence

Item analyses were conducted on all items to obtain internal consistency estimates of reliability for the two subscales. Since the original 67-item SMSCT had only a reliability coefficient of .50 for the assessment subscore and a negative alpha coefficient of −.07 for the intervention subscore, we first looked at the intercorrelation matrix to identify poorly performing items. Next, we employed a stepwise reliability analysis to exclude, one by one, all of the variables having a negative impact on reliability. The item with the largest negative correlation with the subscore was excluded and a new alpha coefficient calculated. This process was repeated until there were no negative



**TABLE 2** Demographic summary by group

| Variable | Scoring system developers ($n$ = 15) | External validity subgroups: Validation subjects ($n$ = 91) | SM group ($n$ = 15) | Ortho group ($n$ = 10) |
|---|---|---|---|---|
| Age | 40.7 (7.0) | 42.3 (10.4) | 47.9 (8.1) | 36.4 (5.4) |
| Gender | 2 males | 18 males | 3 males | 4 males |
| | 13 females | 73 females | 12 females | 6 females |
| Years of clinical practice | 15.1 (7.7) | 15.9 (11.1) | 24.7 (8.2) | 9.2 (3.9) |
| SM service provision | 12.2 (7.6) | 9.0 (9.3) | 20.6 (5.8) | 0.1 (0) |
| Hours/week SM | 22.0 (15.4) | 12.3 (13.9) | 26.2 (12.9) | 0.0 (0) |
| SCI/year | 405.6(278.8) | 34.1 (73.0) | 72.4 (83.7) | 0.2 (0.0) |
| Professional-level degree | 13 entry level | 56 entry level | 11 entry level | 3 entry leve |
| | 2 advanced level | 35 advanced level | 4 advanced level | l7 advanced level |

*Note*. Values are means and SDs.

**TABLE 3** Correlations of items: convergent and discriminate validity evidence

| | Domain | | | Domain | |
|---|---|---|---|---|---|
| Assessment item | Assessment | Intervention | Intervention item | Assessment | Intervention |
| Q1 | 0.33 | –0.01 | Q34 | 0.04 | 0.58 |
| Q2 | 0.57 | –0.19 | Q35 | 0.12 | 0.16 |
| Q3[a] | –0.05[b] | 0.16 | Q37 | 0.12 | 0.17 |
| Q4 | 0.37 | –0.03 | Q38 | –0.08 | 0.30 |
| Q5 | 0.43 | –0.41 | Q39[a] | –0.21 | 0.01 |
| Q7 | 0.30 | –0.20 | Q40 | –0.14 | 0.55 |
| Q8 | 0.48 | –0.07 | Q41 | 0.66 | 0.10[b] |
| Q11 | 0.31 | 0.19 | Q42 | –0.15 | 0.20 |
| Q12 | 0.12 | 0.02 | Q44 | 0.46 | 0.20[b] |
| Q13 | 0.49 | –0.11 | Q45 | 0.26 | 0.26 |
| Q15 | 0.62 | –0.38 | Q46 | 0.07 | 0.48 |
| Q17[a] | 0.08 | –0.03 | Q47 | –0.15 | 0.39 |
| Q19 | 0.35[b] | 0.40 | Q49 | –0.38 | 0.34 |
| Q21 | 0.11 | –0.05 | Q53 | –0.02 | 0.45 |
| Q24 | 0.38 | –0.20 | Q54[a] | –0.07 | 0.04 |
| Q27 | 0.39 | 0.09 | Q55 | –0.33 | 0.48 |
| Q28 | 0.84 | 0.43 | Q58 | –0.38 | 0.61 |
| Q29 | 0.37 | 0.03 | Q64 | 0.00 | 0.15 |
| Q32 | 0.31 | –0.20 | Q66 | 0.26 | 0.07[b] |
| Q33 | 0.13[b] | 0.37 | Q67 | –0.28 | 0.22 |

*Note.* $n$ =15 expert SCI clinicians.
[a]Item performing less than optimally, negligibly correlating with either subtest.
[b]Item more highly correlated with the opposing knowledge scale versus its own.



correlations and only negligible changes in reliability resulted from excluding other items. In this fashion, the selection of the 40 "best" items resulted in Cronbach alpha coefficients for the assessment and intervention subscales of .78 and .71, respectively. These 40 items are listed in Table 3.

To assess the convergent and discriminant validity of the two subscales, we again correlated each item with its own scale (with the item removed) and with the other subscale. The results of these analyses are shown in Table 3. In six cases (Questions 3, 19, 33, 41, 44, and 66), items were more highly correlated with the opposing subscale than with their own. Five questions (12, 21, 35, 37, and 64) had low correlations to their own subscale, and Question 45 had the same correlation for both domains. A total of four items (3, 17, 39, and 54) were found to be performing less than optimally, negligibly correlating with either subscale.

## External Structure Evidence

Independent-samples $t$ tests were used to evaluate the hypothesis that the SMSCT subscores (assessment and intervention) can differentiate between SM experts and orthopedic experts. No significance was found between groups for the assessment subtest ($t = -0.3$, $p = .77$). However, on the intervention subtest, subjects in the Ortho group (49.2 ± 9.2) scored lower than those in the SM group (58.6 ± 11.1). The mean difference was significant ($t = -2.2$, $p = .04$). Figure 4 shows the distributions for the SM and Ortho groups.

Bivariate correlational analyses showed no significant relationships between assessment subscores and predictor variables of clinical expertise. The stepwise multiple linear regression analysis was, therefore, only completed using the intervention criterion variable. Just one variable was identified as a predictor of intervention score. This variable, SM hours per week, accounted for 11% of the variability of the intervention score, $F(1, 87) = 10.62$, $p = .002$, $r^2 = .11$.

## DISCUSSION

To provide valid score interpretation, the technical quality of a test needs to be demonstrated. Results for the SMSCT indicate some significant and noteworthy

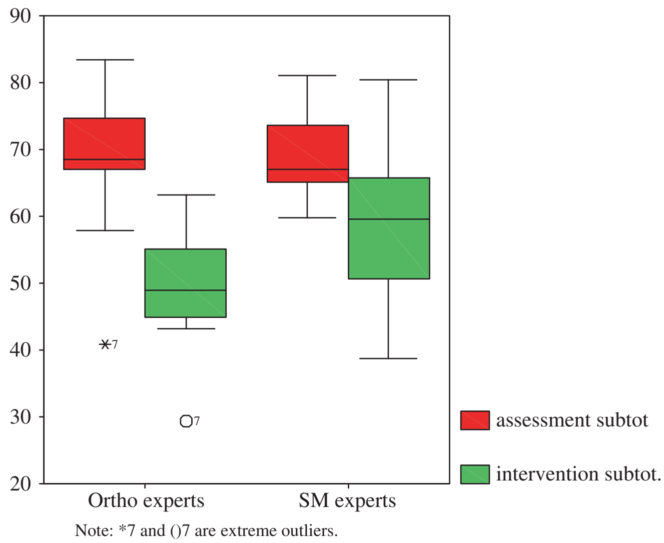


**FIGURE 4** **Box plot of subtest scores by criterion group.**



findings in addition to areas in need of further development. The intercorrelation matrix results showed reasonable item performance, with some items performing better than others. The initial pool of 67 SMSCT items had a reliability coefficient of only .50 for the assessment subtest and −.07 for the intervention subtest. Test reliability was improved to satisfactory levels by removing 27 poorly performing items, resulting in reliability coefficients of .78 and .71, respectively. Since reliability is a measure of item consistency, it is understood that low reliability indicates item inconsistency. Results indicate that a number of poorly functioning items may be attributed to factors such as difficulty with item clarity, terminology, vignette information, or hypotheses. Although item reliability was increased to satisfactory levels by removing poorly performing items, the tradeoff was decreased coverage of content identified as key to representing the population of individuals with SCIs served by SM clinicians (Cohen et al., 2005). The breadth and scope of content covered was therefore narrowed. Clearly, several test items did not perform as expected and therefore will require further revision to improve overall test and individual item performance. The addition of new items and revised items will bring content coverage into alignment with the original test blueprint and defined test content.

The second analysis, in combination with previously obtained content validity evidence, provides some degree of support that the SMSCT measures two unique constructs: knowledge and intervention. We anticipated high correlations for items within a subtest and lower correlations for items across subtests (assessment and intervention). The results indicated that, on the whole, this was the case. Six items, however, were more highly correlated with the opposing scale. Closer inspection of these items did not reveal an apparent reason for this contrary finding. Removal of the six items failed to impact the reliability scores (Cronbach alpha coefficients), and because these items were deemed important to assessment, they were retained. Furthermore, four items were found to contribute little to the overall subtest reliability. Visual inspection of these items did not reveal an obvious explanation, such as content relevance; thus, we decided that these four items should be rewritten prior to future administrations of the tool. For the purposes of these analyses, these problematic items were retained. Another potential limitation of our results is that the values obtained may overestimate the population reliability (alphas) because the sample that was used to create the scoring system was used for the item analyses and the reliability estimates. Therefore, we recommend future research to validate the scoring system and to obtain item and test reliability statistics with unique but comparable groups of SCI experts.

In addition to the technical quality of the test, different sources of evidence are needed to support the meaning and interpretation of a test score. One type of evidence, external structure evidence, measures the extent to which scores converge or diverge with known qualities in the manner expected and is critical to the use and interpretation of the test (American Education Research Association, American Psychological Association, and National Council on Measurement in Education, 1999; Nitko, 2001). To provide evidence that the SMSCT differentiates between two groups with known qualities (the SM and Ortho groups), we hypothesized that if the SMSCT is a measure of clinical expertise, we would expect the scores of SM experts to be higher than those of orthopedic experts. Although assessment and intervention knowledge can both be fundamental facets of clinical expertise that are unique to a given type of expert, our results showed that there were no differences between the groups. The reason could be that the assessment items are not reflective of unique knowledge and skills specific to SM for SCIs, or assessment knowledge is simply not distinct from traditional physical therapy assessment. A significant difference was detected between groups for the intervention subtest, suggesting that these items were capable of detecting differences in intervention knowledge between two groups with known differences.

Another potential limitation of the SMSCT is construct underrepresentation, which refers to the degree to which a test fails to capture important aspects of the construct it purports to measure (Charlin et al., 1998; American Education Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). It implies a narrowed meaning of test scores because the test does not adequately sample the domain of interest. For example, the SMSCT may underrepresent the content domain of assessment knowledge because it does not contain a sufficient variety of skills and knowledge specific to SM for SCIs, including the spectrum of clinical vignettes, hypotheses, and common misconceptions, and because of the elimination of 39% of the assessment items (as previously described).

Finally, to explore which proxy measures of clinical expertise can predict SMSCT scores, we studied characteristics that were hypothesized to affect clinical expertise. We predicted that clinicians who saw many patients with SCIs each year, provided more hours of SM services, worked more years providing SM services, and had advanced-level professional degrees would have higher SMSCT scores on a test specific to the content domain of SM for SCIs. The regression results suggested that clinicians who work more hours per week providing SM services are more likely to have higher intervention subscale scores. However, although significant, SM hours per week only made a small contribution, suggesting that perhaps there are alternate factors that may be stronger indicators of clinical expertise. SMSCT subscores were not related to hypothesized criterion variables of clinical expertise. It is unclear why the anticipated associations were not found. Given this, it should be acknowledged that the SMSCT, in its current form, is not sensitive in detecting differences among chosen experts based on these proxy measures. In the field of SM, there exists no reliable and valid measure of clinical competence or expertise. We therefore had to rely on less precise alternative indicators of expertise. In this work, it is evident that the proxy measures of expertise selected were lacking. It is quite possible that SM expertise involves something more than the characteristics we hypothesized.

By design, we chose to explore associations between SMSCT scores and hypothesized measures of clinical expertise. However, previous research in SCT validation in the area of medicine employed comparison groups including faculty, residents, and clerks as criterion groups for obtaining external structure evidence (Charlin et al., 2000). In order to closer approximate the methodology used in the field of medicine, future studies might employ comparison groups including assistive technology practitioners, therapy interns who have completed all professional coursework, and therapy students. The use of comparison groups with known differences would allow the exploration of potential causal relationships. Exploring the data in this manner would eliminate the impact that occurs from the use of untested associations. Nonetheless, it is clear that more work is needed to identify accurate predictors of clinical expertise.

There is concurrence in the findings indicating that further SMSCT item revision is needed. More vignettes and additional hypotheses incorporating common misconceptions should be considered for inclusion. Pilot test administrations with various groups with known qualities would likely identify whether the SMSCT is comprehensively measuring assessment and intervention knowledge through a wider range of abilities. Further research is needed to establish if perhaps alternate domains (other than assessment and intervention) are more useful as measures of SM expertise for SCIs. Nevertheless, we genuinely believe in the SMSCT item format and conceptual theory underlying the test design as a promising approach to measuring clinical expertise.

## CONCLUSION

Preliminary validation of the SMSCT suggests that the test may be a promising measure of clinical expertise. Additional research should include further development, revision, and testing of items and pilot testing specific to construct representation and item performance. Additional validation is recommended using preestablished groups with known qualities. Future work is planned for the SMSCT with SCIs as well as the development of similar tests in other content domains.

## ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Veterans Affairs through the Center of Excellence for Wheelchairs and Related Technologies (F2181C) and the U.S. Department of Education through the University of Pittsburgh Model Center for Spinal Cord Injury (H133A011107). All work was completed at the Human Engineering Research Laboratories, Pittsburgh, Pennsylvania. We thank the conference coordinators of RESNA and the International Seating Symposium for their assistance with subject recruitment and testing facilities, Anthony Delitto and Sara Piva for assistance with orthopedic expert recruitment, and the clinicians who volunteered their time to participate in this project.

## REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Beaulieu, L. (2007). CMS does an about-face on ATP requirement. *HME Newswire*, 13. Retrieved from http://www.hmenews.com/index.php?p=article&id=hm200712yq7e4d

Beausoleil, S., & van der Vleuten, C. P. (2001). Measurement of clinical reflective capacity early in training as a predictor of clinical reasoning performance at the end of residency: An experimental study on the Script Concordance Test. *Medical Education*, *35*, 430–436.

Charlin, B., Brailovsky, C. A., Brazeau-Lamontagne, L., Samson, L., Leduc, C., & van der Vleuten, C. P. (1998). Script questionnaires: Their use for assessment of diagnostic knowledge in radiology. *Medical Teacher*, *20*, 567–571.

Charlin, B., Desaulniers, M., Gagnon, R., Blouin, D., & van der Vleuten, C. P. (2002). Comparison of an aggregate scoring method with a consensus scoring method in a measure of clinical reasoning capacity. *Teaching and Learning in Medicine*, *14*, 150–156.

Charlin, B., Roy, L., Brailovsky, C., & van der Vleuten, C. P. (2000). The Script Concordance Test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine*, *12*, 189–195.

Chi, M. T. H., & Glaser, R(1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, L. J., Fitzgerald, S., Lane, S., & Boninger, M. (2005). Development of the Seating and Mobility Script Concordance Test (SMSCT) for spinal cord injury: Obtaining content validity evidence. *Assistive Technology*, *17*, 122–131.

Cohen, L. J., Fitzgerald, S. F., Trefler, E., Boninger, M. L., & McCue, M. (2002). Development and reliability testing of a clinical rationale measure of seating and wheeled mobility prescription. In R. Simpson (Ed.), *RESNA 25th International Conference proceedings*. Arlington, VA: RESNA Press.

Davis, D. A., Thomson, M. A., Oxman, A. D., & Haynes, R. B. (1992). Evidence for the effectiveness of CME: A review of 50 randomized controlled trials. *Journal of the American Medical Association*, *268*, 1111–1117.

Fifield, M. G., & Fifield, M. B. (1997). Education and training individuals involved in delivery of assistive technology devices. *Technology and Disability*, *6*, 77–88.

Grossman, J. (1998). Continuing competence in the health professions. *American Journal of Occupational Therapy*, *52*, 709–715.

Herman, J. H., & Lange, M. L. (1999). Seating and positioning to manage spasticity after brain injury. *NeuroRehabilitation*, *12*, 105–117.

Jensen, G. M., Gwyer, J., Shepard, K. F., & Hack, L. M. (2000). Expert practice in physical therapy. *Physical Therapy*, *80*, 28–43.

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Saddle River, NJ: Prentice-Hall.

Norman, G. (2000). The epistemology of clinical reasoning: Perspectives from philosophy, psychology, and neuroscience. *Academic Medicine*, *75*, 127–133.

Schmidt, H. G., Norman, G., & Boshuizen, H. (1990). A cognitive perspective on medical expertise: Theory and implication. *Academic Medicine*, *65*, 611–621.

Tardif, J., & Boshuizen, H. P. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine*, *75*, 182–190.