

Original Research

# Novice Inter-Rater Reliability on the Selective Functional Movement Assessment (SFMA) After a 4-Hour Training Session

Brent Harper<sup>1</sup><sup>a</sup>, Adrian Aron<sup>2</sup><sup>1</sup> Physical Therapy, Chapman University, <sup>2</sup> Physical Therapy, Radford University

Keywords: interrater reliability, movement, education, compensation patterns, students, teaching

<https://doi.org/10.26603/001c.82173>

---

## International Journal of Sports Physical Therapy

Vol. 18, Issue 4, 2023

---

### Background

The Selective Functional Movement Assessment (SFMA) is a whole-body movement classification system that identifies non-optimal movement performance requiring further assessment. There needs to be more evidence specifying the training time required to obtain SFMA reliability for entry-level health care practitioners.

### Purpose

The primary intent of this study was to determine SFMA inter-rater reliability between two third-year physical therapy students following an in-person three-hour training and one-hour follow-up training with a certified SFMA physical therapist. The secondary purpose was to compare rater scores of the composite criterion 50-point checklist and rater categorization using the top-tier movements in real-time assessments of healthy participants.

### Study Design

Inter-rater reliability study.

### Methods

Two novice raters received training on assessing movement using the SFMA. Participants included non-pregnant healthy adults screened for general exercise, participants were excluded for history of orthopedic surgery within the prior six months. Three independent raters, including two novice and one SFMA-certified rater, individually assessed the top-tier movements in separate rooms in real-time. Participants were randomly assigned a start location, and raters were blinded to each other's criterion 50-point checklist and categorical scoring. Statistical analysis included a paired *t*-test, a repeated measures ANOVA, and a two-way, mixed absolute agreement ICC.

### Results

Twenty-five participants (23.4 years  $\pm$  1.9; 72% female) completed the SFMA top-tier movements. Significant differences were identified with novice raters identifying fewer non-optimal movement patterns than the certified clinician. The intraclass correlation coefficient (ICC<sub>2,1</sub>) was moderate (0.60, *p*<0.001) for all three raters on the 50-point criterion checklist scoring.

---

a Corresponding Author:

Brent A. Harper  
Chapman University, Department of Physical Therapy,  
9401 Jeronimo Road, Irvine, CA, 92618, USA  
E-mail: [brharper@chapman.edu](mailto:brharper@chapman.edu)  
Phone: 714-516-5946  
Fax: 949-206-0012  
ORCID: 0000-0002-4705-5653

## Conclusion

Third year physical therapy students were able to demonstrate moderate inter-rater reliability assessing healthy individuals using the 50-point criterion checklist. Variation between novice raters may reflect the amount of previous exposure assessing movement and suggests that some may require more time learning and practicing in order to identify non-optimal movement patterns that may require further assessment.

## Level of Evidence

3b

©The Author(s)

## INTRODUCTION

Musculoskeletal screening and movement-based assessment tests are used in physical therapy and sports medicine to identify injury risk, movement dysfunction, and sources of pain.<sup>1,2</sup> Screening and assessment tests incorporating whole body kinetic chain movements may reveal underlying impairments contributing to a person's chief complaint of pain in regions elsewhere in the body.<sup>1</sup> Often, dysfunctional movement patterns and impaired motor control are exhibited in response to pain or limited joint motion.<sup>3</sup> When aberrant movement patterns are present, other systems in the body compensate to complete functional movement patterns of daily life.<sup>4-6</sup> These compensated functional movements are considered dysfunctions.<sup>2-5</sup>

Regional interdependence (RI) is the theoretical construct which proposes that adjacent anatomical areas may contribute to or be the primary source of musculoskeletal symptoms. Thus, regardless of proximity to the anatomical site of symptoms, non-symptomatic dysfunction(s) in various body regions within the kinetic chain may directly or indirectly influence the clinical presentation.<sup>1,2,7-9</sup> The Selective Functional Movement Assessment (SFMA) is based on RI.<sup>10,11</sup> The pathoanatomical model may lead to misdiagnoses for the source of pain; in contrast, RI provides a more comprehensive approach to identify and assess multiple relevant body regions for their possible role in the clinical presentation.<sup>9-11</sup> From the RI model, health care providers can analyze movements and postures and identify the specific movements that can lead to abnormal loads or abnormal pressure and adaptive changes resulting in changed kinematics.<sup>12</sup>

## CLASSIFICATION SYSTEMS

The SFMA is a musculoskeletal screening and diagnostic classification tool in which top-tier movements are scored at two levels. First, the composite criterion checklist is scored out of 50, then movements are assigned to a categorical identifier, filtering the movement into one of four categories.<sup>1,2,7,8</sup> The SFMA begins with a foundational movement evaluation to identify impairments and limitations within tri-planar movements and to determine whether or not those movements provoke patient symptoms.<sup>1,2,11</sup> Tri-planar movements are scored using a composite checklist and then each pattern is categorized into one of four categories: functional non-painful (FN), functional painful (FP), dysfunctional non-painful (DN), or dysfunctional

painful (DP). Movements scored as DN are further assessed for mobility or motor control impairments contributing to the original site of pain and dysfunction.<sup>1,2,7,8,11</sup>

Understanding the inter-rater reliability and validity of screening and diagnostic tools is an essential factor to consider before clinical utilization.<sup>13,14</sup> The inter-rater reliability of the SFMA is not established for healthcare practitioners in training and with real-time assessment (a researcher physically evaluating the participant, one-on-one); thus, the substantiality of the test is still unknown.<sup>13</sup>

To date, three studies<sup>1,7,8</sup> have evaluated the reliability of the top-tier SFMA. In 2014, Glaws et al.<sup>1</sup> performed top-tier SFMA inter-rater and intra-rater video assessments on healthy subjects using three SFMA-certified raters with varying levels of SFMA clinical application. In 2017, Dolbeer et al.<sup>8</sup> assessed inter-rater reliability on subjects with pain with two raters assessing in real-time and one rater scoring by watching a video recording; however, all raters were SFMA-certified with over 400 hours of clinical application using this method. In 2019, Stanek et al.<sup>7</sup> evaluated inter-rater and intra-rater reliability on healthy subjects using three raters in real-time comparing two SFMA-certified raters with varying levels of SFMA clinical application while one rater was a student with no formal SFMA top-tier training but who did have a summer clinical rotation utilizing this approach.

## SIGNIFICANCE OF THE PROBLEM

Although prior research<sup>1,7,8</sup> has demonstrated good reliability, healthcare practitioners in training have not been thoroughly studied, nor has the standard training time required to develop reliable movement pattern assessment and recognition. When two researchers are scoring the same participant simultaneously, only one can give the instruction to the patient, which, in itself, limits the inter-rater reliability results.<sup>13</sup> Assessing movement live involves different viewing positions than a video assessment. Furthermore, watching video performance removes the real world setting in which clinical application occurs. To avoid these limitations, researchers need to individually perform the SFMA. The primary intent of this study was to determine SFMA inter-rater reliability between two third-year physical therapy students following an in-person three-hour training and one-hour follow-up training with a certified SFMA physical therapist, and the secondary purpose was to compare rater scores of the composite criterion 50-point checklist and rater categorization using the top-

**Table 1. Subjects' Descriptive Statistics**

	Female	Male	Total
Number of Participants	18	7	25
Age (years)	23.2 ± .8	24.0 ± .7	23.4 ± 1.9
Height (cm)	164.7 ± 1.4	175.2 ± 2.5	167.6 ± 7.6
Weight (kg)	65.1 ± 2.9	78.5 ± 1.8	68.9 ± 12.3
BMI	23.9 ± 1.0	25.4 ± .7	24.4 ± 4.0

Note. Abbreviations: cm, centimeters; Kg, kilograms; BMI, Body Mass Index; Values are presented as Mean ± SD

tier movements in real-time assessments of healthy participants.

## METHODS

### PARTICIPANTS

A convenience sample of 29 healthy volunteers were recruited. Participants were excluded if they had any positive marks on the Physical Activity Readiness Questionnaire (PAR-Q) health assessment, had undergone orthopedic surgery within the prior six months, were currently pregnant or thought they might be pregnant, or were under the age of 18. Four participants reported pain during the testing procedures and were excluded to align with the inclusion criterion of healthy subjects and to mitigate error by maintaining a homogenous sample and similar to Glaws et al.<sup>1</sup> and Stanek et al.<sup>7</sup> which both included healthy participants without reported pain. The study had IRB approval. All subjects signed informed consent. The final sample consisted of 25 participants (7 male, 18 female). Participant demographics are provided in [Table 1](#).

### STUDY DESIGN

#### EQUIPMENT/ MATERIALS USED DURING DATA COLLECTION

Participants completed the top-tier SFMA movement patterns administered with verbal instruction, live and individually, by each rater (Appendix A). The standardized one-page scoring sheet was used to eliminate any discrepancy in criteria interpretation between researchers (Appendix B). No additional scripts, surveys, or specific software/equipment were used to conduct this study.

#### SFMA ADMINISTRATION TRAINING

The two student raters, rater 1 and rater 2, were third-year doctor of physical therapy students with no clinical experience performing the top-tier SFMA. Rater 1 completed an athletic training program but was not licensed and had never practiced as an athletic trainer. They participated in a three-hour in-person training course on conducting and scoring the SFMA with a certified SFMA physical therapist. The certified SFMA physical therapist serving as rater 3 had 15-hours of formal SFMA training, three-years of clinical application, and 18-years of overall clinical practice. The

certified SFMA physical therapist provided the training using the SFMA training videos as the standard of instruction for teaching the top-tier movements. The instructor also demonstrated in-person to the raters the appropriate procedures for testing including movements, appropriate cues, and the importance of using multiple planes of view in order to fully appreciate the movement demonstrated. After the training and before data collection, a pilot study for internal validity was conducted on three participants concurrently assessed using the top-tier movements by all three raters. All raters recorded their scores on the participants' movements and interpreted the results, identifying and discussing any areas of discrepancy between the raters. Variance was found during the scoring of the upper extremity and multi-segmental rotation movements. One hour of additional in-person training was administered in order to increase scoring consistency between the raters, resulting in a total of four hours of training.

#### TESTING PROCEDURE

The participants entered a room in the testing facility and were given a participant number. Participants completed the informed consent and HIPAA form and were screened for inclusion and exclusion criterion prior to testing. The researcher reviewed these forms with the participants and answered any questions. Participants then completed and signed consent forms.

Participants were randomly assigned to one of three examination rooms and then rotated through the other two rooms. Upon entrance, the administering rater would direct the participant to complete each of the desired top-tier movements of the SFMA. Each rater assessed the participants live and individually in separate rooms. After a brief demonstration of the desired motion, the rater assessed and scored movements as participants performed the SFMA top-tier movements. The participants were allowed three attempts to perform each specific movement, and the rater could observe the movements from any direction in order to obtain multiple planes of view.

Each participant was instructed to perform the top-tier movements in the specific order on the scoring sheet (Appendix B). This method is standardized through the SFMA in order to select the appropriate primary categorical pattern. Bilateral movements were standardized right side followed by left. After each movement, the participants were asked if they experienced any pain and their results were documented accordingly. The researchers were blinded to each other and each independently scored the SFMA with the 50-point criterion checklist scoring tool and the categorical scoring sheet. The participants were escorted to a different rater upon completion with their first rater in a sequential order after the initial randomization.

#### STATISTICAL ANALYSIS

Composite scores were derived from the 50-point criterion checklist and were compared between researchers. Data analysis was carried out using SPSS Version 27 (IBM Corporation, Armonk, NY). The inter-rater reliability was cal-

culated using intraclass correlation coefficients (ICC). A paired *t*-test assessed the difference in scores between rater 1 and 2 and a repeated measures ANOVA compared scores from raters 1 and 2 with rater 3. A two-way, mixed absolute agreement ICC analyzed results of the composite score to quantitatively measure the reliability and absolute agreement between the researchers. Normality was assessed using the Shapiro-Wilk test and Mauchly's test of sphericity. Partial eta squared effect size was interpreted as small 0.01, 0.06 medium, and 0.14 large.<sup>15</sup> In order to control for type I error, a simple contrast correction was performed. Rater 3 was used as the reference category by which the other raters (e.g., rater 1 and 2) were compared. A Bonferroni correction established a new alpha of  $p=.017$ . ICC values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability.<sup>15,16</sup>

Categorical classification of the top-tier functional movements for FN and DN used Cohen's Kappa to measure the agreement between researchers. Cohen's Kappa coefficient measured reliability between all three raters and to determine the likelihood their agreement was due to chance. Data were interpreted as statistically significant different with a *p*-value less than 0.05. Cohen's Kappa results were categorized as: 0.01-0.2 (1%-20%) representing slight agreement, 0.21-0.40 (20%-40%) representing fair agreement, 0.41-0.60 (40%-60%) corresponding to moderate agreement, 0.61-0.80 (60%-80%) accepted as substantial agreement, and 0.81-0.99 (80%-90%) considered almost perfect agreement, with 1.00 (100%) representing a perfect agreement.<sup>15</sup>

## RESULTS

Twenty-five volunteers, seven male and eighteen females with mean age of  $23.4 \pm 1.9$  years and BMI of  $24.4 \pm 4.0$ , were analyzed and scored, subject demographic information is provided in [Table 1](#). Normality was met.

There were significant differences in the top-tier 50-point criterion checklist between rater 1 and rater 2,  $t(24)=4.594$ ,  $p<0.001$  with a large effect size (Cohen's  $d = 3.831$ ). Rater 1 identified more deviations from optimal top-tier movement performance using the checklist standard (Appendix B) than rater 2 ([Table 2](#)). Thus, rater 1 consistently identified more non-optimal movement patterns requiring further assessment during the 50-point criterion checklist assessment.

There were also significant differences in the 50-point criterion checklist between rater 1 and rater 3,  $F(1,24)=51.059$ ,  $p<0.001$  with large partial eta squared effect size (.680), as well as between rater 2 and rater 3,  $F(1,24)=111.484$ ,  $p<0.001$  with an even larger partial eta squared effect size (.823). Each rater identified a different total number of non-optimal movements out of the 50-points possible. After scoring all 25 participants using the 50-point criterion checklist scores, rater 1 identified a mean of 9.7 and SD of .9 non-optimal movement patterns, rater 2 identified a mean of 6.2 and SD of .7, and rater 3

**Table 2. 50-Point Criterion Checklist Composite Score**

	Rater 1	Rater 2	Rater 3
Composite	9.7 ± .9	6.2 ± .7	13.9 ± .8
Minimum	2	1	8
Maximum	19	15	24

Note. Values are presented as Mean ± SD; Composite score is the number of non-optimal movements identified in the 50-point criterion checklist.

**Table 3. Inter-rater Reliability 50-Point Criterion Checklist Score**

	ICC [2,1]	SEM	MDD
All Raters	0.6	4.1	11.4
Rater 3 to Rater 1	0.78	3.8	10.5
Rater 3 to Rater 2	0.55	3.8	10.5
Rater 1 to Rater 2	0.56	4	11.1

Note. Values are presented as Mean ± SD; SEM=Standard Error of the Mean; MDD=Minimal Detectible Difference

identified a mean of 13.9 and SD of .8 ([Table 2](#)). Rater 1 and rater 2 had significant differences in the 50-point criterion checklist from each other and also when compared to the clinical expert, rater 3. Both rater 1 and 2 identified non-optimal movements from the top-tier screen; however, they did not identify as many when compared to the SFMA certified clinical expert. In fact, rater 1 identified 3.5 more non-optimal movements than rater 2. Rater 3 identified 7.7 more non-optimal movements than rater 2 and 4.2 more than rater 1.

The agreement between researchers for the 50-point criterion checklist results utilizing the intraclass correlation coefficient (ICC<sub>2,1</sub>) measure was 0.60,  $p<0.001$ , demonstrating an overall moderate agreement between raters,  $F(24,48)=6.13$ ,  $p<0.001$  ([Table 3](#)). A higher composite score on the 50-point criterion checklist indicates more non-optimal movement patterns were identified (Appendix B, left column). Rater 3 identified the most non-optimal movement patterns and therefore had a higher scoring mean ( $13.92 \pm 4.17$ ), while rater 2 had the lowest scoring mean ( $6.24 \pm 3.39$ ). Rater 3 and rater 1 had a reliability of 0.78; however, reliability between rater 3 and rater 2 was 0.55 and reliability between rater 1 and rater 2 was 0.56 respectively. Cronbach's alpha was used for internal consistency to examine each individual rater on how well they performed on scoring the 50-point criterion checklist. The item-to-total correlation for rater 2 was 0.59 suggesting that rater 2 was not similar to raters 1 and 3. When rater 2 scores were removed, the reliability increased between raters 3 and 1 with a Cronbach's alpha of 0.87. This indicates that rater 2 was not assessing the movements as accurately as raters 3 and 1. Internal consistency should demonstrate a moderate correlation, somewhere between 0.70 and 0.90 respectively.<sup>17</sup>

Categorical scoring results of the inter-rater reliability between all raters using Cohen's kappa values ranged between slight and moderate depending on the movement pattern. Mean categorical kappa value scores for all raters

**Table 4. Inter-rater Reliability Categorical Classification of Movements**

	<u>Rater 3 to 1</u>	<u>Rater 3 to 2</u>	<u>Rater 1 to 2</u>
	Cohen's Kappa	Cohen's Kappa	Cohen's Kappa
Cervical Flex	.68	.68	.68
Cervical Ext	.30	.12	.36
Cervical R Rot	.36	.04	.16
Cervical L Rot	.30	.05	.01
UE R Pattern 1	.51	.66	.49
UE L Pattern 1	.53	.14	.23
UE R Pattern 2	-.03	-.04	.04
UE L Pattern 2	-.02	.02	.26
MS Flex	.07	.16	.45
MS Ext	---	---	-.06
MS R Rot	.58	-.03	.09
MS L Rot	.13	.01	.14
SLS R	.06	.12	.48
SLS L	.30	.09	.64
Deep Squat	-.03	-.03	.75
Overall Mean	.25	.13	.32
% Agree DNs	0.64	0.63	0.45

Note. Flex: Flexion; Ext: Extension; Rot: Rotation; UE: Upper Extremity; MS: Multi-Segmental; SLS: Single-Leg Stance R=right; L=left. % Agree DNs=Percent Agreement identifying Dysfunctional Non-Painful

was only fair; however, the mean percent agreement in the ability to identify DNs, which is important to identify since they warrant further assessment, was substantial (Table 4).

## DISCUSSION

There is currently limited research available to describe the inter-rater reliability of raters scoring the top-tier SFMA. The primary intent of this study was to determine SFMA inter-rater reliability between two third-year physical therapy students following an in-person three-hour training and one-hour follow-up training with a certified SFMA physical therapist, and the secondary purpose was to compare rater scores of the composite criterion 50-point checklist and rater categorization using the top-tier movements in real-time assessments of healthy participants.

Three studies<sup>1,7,8</sup> have previously evaluated the reliability of the SFMA. For consistency, this study followed the same top-tier SFMA movements (Appendix A) as described by Glaws et al.,<sup>1</sup> Dolbeer et al.,<sup>8</sup> and Stanek et al.<sup>7</sup> Furthermore, the standardized one-page scoring sheet used in this study (Appendix B) was the same presented by Glaws et al.<sup>1</sup> and Dolbeer et al.<sup>8</sup> in their appendices and as described by the Functional Movement Systems, Inc. training materials. For this study for rater convenience when assessing and scoring the movements, the Criterion Checklist and Categorical scoring forms were combined on a single page for easier data completion (Appendix B).

The results of the current study indicate a significant difference with large effect size between raters 1 and 2. On

average, rater 1 identified 3.43 more non-optimal movements than rater 2. The current study differed from prior SFMA reliability studies<sup>1,7,8</sup> in that this study sought to assess if novice raters with only four hours of training and no clinical experience could consistently score in real-time, independently, using the SFMA 50-point criterion checklist. Stanek et al.<sup>7</sup> included an undergraduate athletic training student who had fair reliability compared with two SFMA certified clinicians after completing a clinical rotation utilizing this method. In this study, rater 1 completed an athletic training program prior to directly entering the physical therapy program. It may be that movement assessment practice gained through the athletic training program enabled rater 1 to identify more non-optimal movement than rater 2. Therefore, despite having the same amount of in-person training and practice with the SFMA top-tier prior to study initiation, rater 1 with more movement assessment exposure performed better. Another possible explanation may relate to the additional experience rater 1 may have gained following assessment protocols during the athletic training program, allowing the rater to engage more familiarly with a structured, systematic format such as the SFMA assessment procedure.

The results of the present study demonstrated a significant difference with large effect size between novice raters when compared to the SFMA certified clinician. Although both novice raters identified non-optimal movements, they were not able to identify the same number as the clinical expert. Stanek et al.<sup>7</sup> included a student, but all raters were in the room when SFMA scoring took place and the expert rater provided all verbal instructions, relieving the student



of any responsibility or direct interaction with the participants. In contrast, student raters in this study were alone in the examination rooms and were individually responsible for verbally directing (Appendix A), assessing, and recording results in real-time. Thus, rather than simply observing and recording, these two student raters were responsible for the entire process. This may have increased the cognitive load as the raters attempted to identify as many items on the 50-point criterion checklist while also remembering to observe the quality of the movement (Appendix B, SFMA 50-Point Criterion Checklist). This distraction by the process may explain why student raters had difficulty assessing for and marking specifically articulated non-optimal movement patterns. Therefore, if the students failed to mark non-optimal movement on the criterion checklist, it would lead to inappropriate categorization.

The findings of the current study indicate that the top-tier SFMA inter-rater reliability between the three raters yielded poor to moderate agreement ( $ICC=0.60$ ,  $CI_{95} [.02-.84]$ ,  $p<0.001$ ) of the composite scoring for all subjects (Table 3). Dolbeer et al.<sup>8</sup> found similar inter-rater reliability ( $ICC=0.61$ ,  $CI_{95} [.45-.73]$ ) of the criterion checklist composite score between three certified SFMA raters with over 400 hours of application. The similarity in inter-rater reliability despite more SFMA clinical experience in the Dolbeer et al.<sup>8</sup> study may be due to the fact that their raters assessed individuals with pain while this study focused on healthy participants. The poor to moderate inter-rater reliability ( $ICC=0.43$ ,  $CI_{95} [.12-.67]$ ) of the study by Glaws et al.<sup>1</sup> may have been related to the use of video, rather than live, assessment. Allowing the rater to examine the participant with control over the instructions and the ability to move around the participant, as was done in the current study, may have allowed the student raters to demonstrate a greater degree of inter-rater reliability than the more experienced raters in Glaws et al.<sup>1</sup>

There are a limited number of studies examining the reliability of the top-tier SFMA. Prior research assessed reliably using video and real-time concurrently with varying levels of SFMA training and clinical practice resulting in varying levels of agreement,  $ICC=0.43$ ,  $CI_{95} [.12-.67]$ <sup>1</sup> and  $ICC=0.61$ ,  $CI_{95} [.45-.73]$ ,<sup>8</sup> respectively. The current study obtained similar or better reliability ( $ICC=0.60$ ,  $CI_{95} [.02-.84]$ ,  $p<0.001$ ) from a four-hour educational training of health care providers in their third year of training, with minimal clinical exposure, and no use of SFMA in the clinic. This adds significantly to the research for a health care profession which continually assesses movement in the clinic since it is the first study to assess SFMA reliability in a real-time, clinical situation and in which the raters performed the SFMA evaluation separately.

In summary, individuals in an entry-level physical therapy educational program may not be able to identify all non-optimal movement patterns using the SFMA top-tier after limited training. Inexperience, combined with limited training time, may explain the lack of agreement between the novice and certified raters in the categorical scores. Novice student raters may have an inability to distinguish the degree of effort and asymmetry during the movement

scoring when using the 50-point criterion checklist. The novice raters seemed to be more concrete, focusing on the completion of the movement without evaluating the overall quality of the movement pattern. Thus, the novice might mark a movement performed with excessive effort as complete without marking the excessive effort or lack of motor control. This, in turn, would cause the student rater to classify an effortful or uncontrolled movement as FN rather than DN. This lack of attention to details related to the quality of the movement pattern was noted in the pilot study and was a primary reason for the additional hour of training. Although the hour appeared to be sufficient in the pilot study, this gain was not retained several days later during the data collection. Those with more clinical experience and those who have completed other movement-based education with clinical rotations (i.e., athletic training program) may be more proficient at assessing movement patterns. Although four hours of training may not be sufficient to allow novice practitioners to identify musculoskeletal impairments requiring further clinical assessment, some novice third year doctor of physical therapy student practitioners demonstrate more ability to perform movement assessment, perhaps due to greater exposure to movement assessments, to identify potential clinical regions that require further assessment.

#### LIMITATIONS

Since this study involved live examinations in which the participants were examined three times in a row with a few minutes between each researcher, participants may have experienced a learning effect. However, this effect was controlled through the randomization of the order in which participants were seen by the researchers. Furthermore, the participants were not coached on strategies to improve their movement during the evaluation as the raters were observing the instinctive, non-guided movement patterns of the individuals. Future real-time reliability studies might consider video recordings of each rater's scoring to control for the possibility of subjects modifying or changing categories with a few repetitions of a movement pattern. Since the two novice raters were third year doctor of physical therapy students, as opposed to seasoned clinicians, they had minimal experience with both the SFMA and the evaluation of participant movement patterns within a clinical setting. However, this could be considered an accurate representation of new graduates who want to incorporate the SFMA into their assessments upon initiation of their careers. The target population of this study consisted of healthy participants that produced no positive marks on their PAR-Q form, while the SFMA is designed to be utilized on clinical patients that are experiencing musculoskeletal pain.

#### CONCLUSION

The results of this study indicate third year physical therapy students were able to demonstrate moderate inter-rater reliability in assessing healthy individuals using the

50-point criterion checklist for top-tier SFMA. There were differences between student raters. Variation between novice raters may reflect the amount of time accrued assessing movement and suggests that some students may require more time learning the steps involved and practicing movement assessment in order to identify non-optimal movement patterns that may require further assessment.

.....

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGMENTS

Authors would like to thank Daniel Harrell, Zachary Harrelson, Chelsea Hermann, and Alison Phillips with their assistance in this research project and Sonya Harper for editing assistance.

SOURCE OF FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Submitted: March 17, 2023 CDT, Accepted: May 16, 2023 CDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

## REFERENCES

1. Glaws KR, Juneau CM, Becker LC, Di Stasi SL, Hewett TE. Intra- and inter-rater reliability of the selective functional movement assessment (sfma). *Int J Sports Phys Ther.* 2014;9(2):195-207.
2. Fauntroy V, Fyock M, Hansen-Honeycutt J, Nolton E, Ambegaonkar JP. Using the selective functional movement assessment for the evaluation of dancers' functional limitations and dysfunctions: a critically appraised topic. *J Sport Rehabil.* 2019;28(8):891-896. [doi:10.1123/jsr.2018-0054](https://doi.org/10.1123/jsr.2018-0054)
3. Lund JP, Donga R, Widmer CG, Stohler CS. The pain-adaptation model: a discussion of the relationship between chronic musculoskeletal pain and motor activity. *Can J Physiol Pharmacol.* 1991;69(5):683-694. [doi:10.1139/y91-102](https://doi.org/10.1139/y91-102)
4. Shumway-Cook AW, M.H. *Motor Control: Translating Research into Clinical Practice.* Lippincott Williams & Watkins; 2012.
5. Rhinehart AJ, Schroeder KM, May J, Baker R, Nasypany A. Movement assessment: Techniques and possible integration into clinical practice. *Int J Athl Ther Train.* 2015;20(6):5-9. [doi:10.1123/ijatt.2015-0021](https://doi.org/10.1123/ijatt.2015-0021)
6. Hodges PW, Tucker K. Moving differently in pain: a new theory to explain the adaptation to pain. *Pain.* 2011;152(3 Suppl):S90-S98. [doi:10.1016/j.pain.2010.10.020](https://doi.org/10.1016/j.pain.2010.10.020)
7. Stanek JM, Smith J, Petrie J. Intra- and inter-rater reliability of the selective functional movement assessment (sfma) in healthy participants. *Int J Sports Phys Ther.* 2019;14(1):107-116. [doi:10.26603/ijstpt20190107](https://doi.org/10.26603/ijstpt20190107)
8. Dolbeer J, Mason J, Morris J, Crowell M, Goss D. Inter-rater reliability of the selective functional movement Assessment (sfma) by sfma certified physical therapists with similar clinical and rating experience. *Int J Sports Phys Ther.* 2017;12(5):752-763. [doi:10.26603/ijstpt20170752](https://doi.org/10.26603/ijstpt20170752)
9. Cheatham SH, D., Kreiswirth E. The regional interdependence model: A clinical examination concept. *Int J Athl Ther Train.* 2014;19(3):8-14.
10. Wainner RS, Whitman JM, Cleland JA, Flynn TW. Regional interdependence: a musculoskeletal examination model whose time has come. *J Orthop Sports Phys Ther.* 2007;37(11):658-660. [doi:10.2519/jospt.2007.0110](https://doi.org/10.2519/jospt.2007.0110)
11. Cook G. *Movement: Functional Movement Systems: Screening, Assessment and Corrective Strategies.* Target Publications; 2010.
12. Miller RCH, W.J., Yound IA, Klausner SH. Application of regional interdependence in a 20 year old male collegiate baseball player with recurrent non-specific low back pain: A retrospective case report. *Ortho Phy Ther Prac.* 2012;21(1):14-20.
13. Monnier A, Heuer J, Norman K, Äng BO. Inter- and intra-observer reliability of clinical movement-control tests for marines. *BMC Musculoskelet Disord.* 2012;13(1):263. [doi:10.1186/1471-2474-13-263](https://doi.org/10.1186/1471-2474-13-263)
14. Harper BA, Glass SM. Item-level and composite-level interrater reliability of functional movement screen scores following condensed training in novice raters. *Int J Sports Phys Ther.* 2021;16(4):1016-1024. [doi:10.26603/001c.25793](https://doi.org/10.26603/001c.25793)
15. Portney LGW, M.P. *Foundations of Clinical Research.* Vol 3rd. Pearson Education, Inc.; 2009.
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155-163. [doi:10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012)
17. Thomas RL, Zidan MA, Slovis TL. What you need to know about statistics Part I: validity of diagnostic and screening tests. *Pediatr Radiol.* 2015;45(2):146-152. [doi:10.1007/s00247-014-2882-7](https://doi.org/10.1007/s00247-014-2882-7)



## SUPPLEMENTARY MATERIALS

### **Appendix A**

Download: <https://ijspt.scholasticahq.com/article/82173-novice-inter-rater-reliability-on-the-selective-functional-movement-assessment-sfma-after-a-4-hour-training-session/attachment/170314.pdf>

---

### **Appendix B**

Download: <https://ijspt.scholasticahq.com/article/82173-novice-inter-rater-reliability-on-the-selective-functional-movement-assessment-sfma-after-a-4-hour-training-session/attachment/170315.pdf>

---