

Establishing Intra- and Inter- tester Reliability of the Sustained Athlete Fitness Exam in Competitive Older Athletes

ALEXIS BUTTERFIELD, SPT, ALEXA HENRY, SPT, SYDNEY JENSEN, SPT, ALEXA STEFFL, SPT, ASHLYN O'FARRELL
RESEARCH ADVISOR: DR. BECCA JORDRE

Background

The Sustained Athlete Fitness Exam (SAFE) is a functional assessment tool incorporating different subscales including cardiovascular (CFIT), muscular (MFIT), flexibility (FFIT), and balance (BFIT) fitness measures with age- and sex-specific norms derived from National Senior Games athletes. While the individual tests comprising the SAFE demonstrate good-to-excellent reliability in isolation, the inter-rater and intra-rater reliability of the SAFE as a complete assessment tool has not been established.

Purpose

The purpose of this study was to evaluate the inter-rater and intra-rater reliability of the SAFE test and subscales in older adult athletes participating in the National Senior Games.

Study Design

Repeated-measures observational reliability study



Methods

Participants

- N=66
- Age 65.9 ± 8.0, range 52-80 years old
- Male = 28, Female = 38
- Registered in the 2025 National Senior Games, Des Moines, IA
- ≥50y/o

Raters

- 2 second-year Student Physical Therapists (SPT)
- 2 licensed physical therapists (PT) (novice & experience)
- Raters completed standardized online SAFE training course with video instruction & competency assessments prior to data collection

Design

- Testing conducted over 6 days at National Senior Games

Reliability Analysis

- Inter-Rater: Raters simultaneously & independently scored each participant on site.
- Intra-Rater: Sessions were video/photo recorded; same two raters re-scored 6 months after initial scoring.

Data Analysis

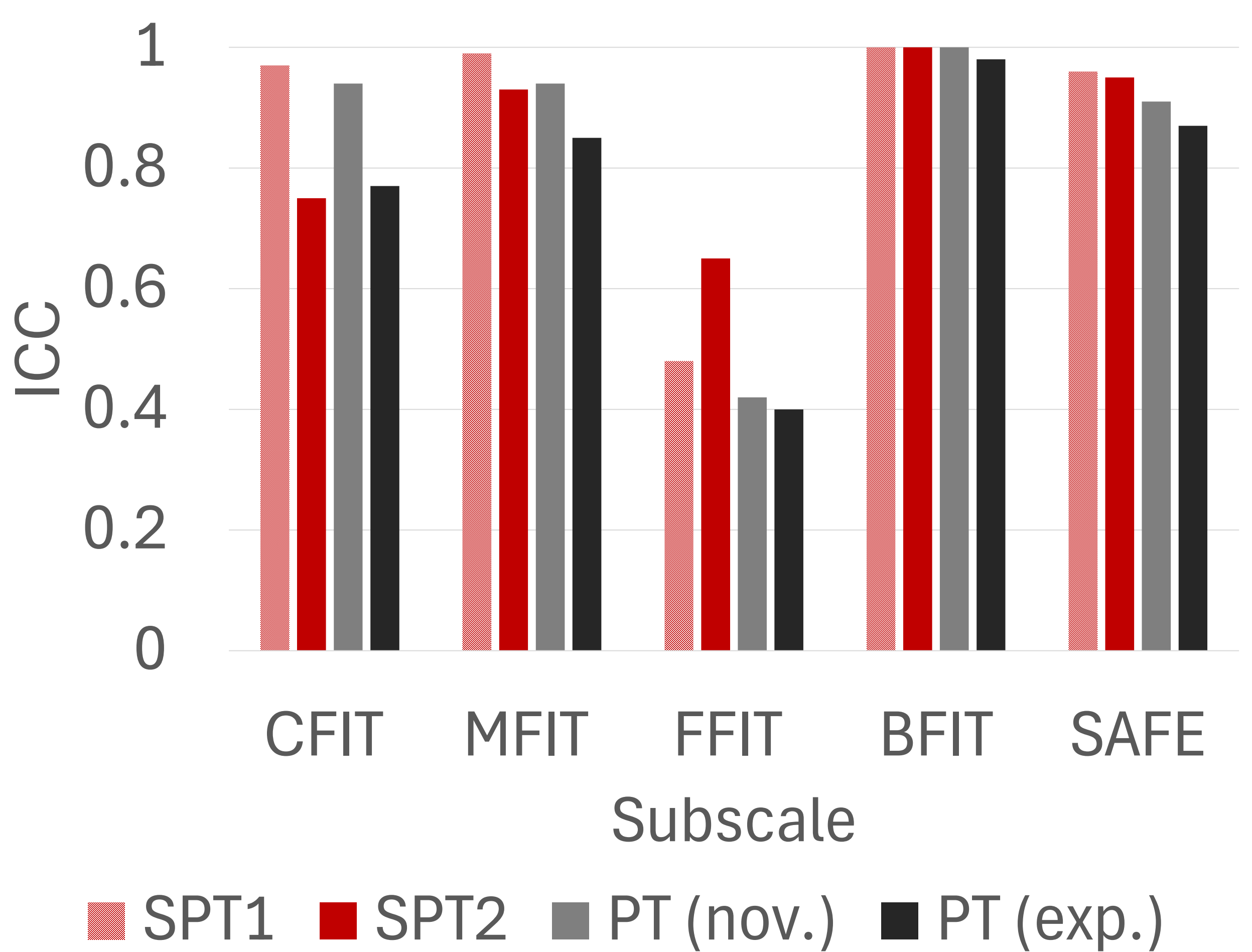
- SPSS 31.0
- Descriptive statistics
- ICCs from one-way random-effects model
- Standard error of measurement (SEM) & Minimal detectable change at 95% confidence (MDC₉₅) derived from paired rating differences.
- Alpha < 0.05

Results

Inter-rater Reliability (ICC)

	ICC (95% CI)	Classification
CFIT	0.78 (0.67 - 0.86)	Good
MFIT	0.98 (0.97-0.99)	Excellent
FFIT	0.62 (0.45-0.75)	Moderate
BFIT	0.97 (0.95-0.98)	Excellent
SAFE	0.94 (0.91-0.96)	Excellent

Intra-rater Reliability (ICC)



	SEM	MDC ₉₅
CFIT	0.54	1.5
MFIT	0.26	0.73
FFIT	0.48	1.32
BFIT	0.25	0.69
SAFE	0.83	2.3

Conclusion

- The composite SAFE demonstrates excellent inter-rater and good intra-rater reliability.
- Subscales of the SAFE have moderate-excellent inter-rater and **good to excellent** intra-rater reliability with the exception of the FFIT subscale which demonstrates poor to moderate intrarater reliability
- The scale error of the SAFE is low
- A difference in scores of 2-3 points is needed to see a clinically meaningful change

Limitations

The small sample of 2 PTs and 2 SPTs may limit generalizability of these outcomes along with the environmental distractions of collecting interrater reliability in a field vs. laboratory setting. Lower FFIT reliability may have been influenced by study design with changes in range of motion related to long holds for dual testing while intra-rater retest methods utilizing video and photograph may have introduced variability not attributable to true scoring differences.

Clinical Relevance

The SAFE composite score is a reliable measure for assessing and monitoring fitness in older athletes, with clinically meaningful change defined as greater than or equal to 2-3 points.

Reference Link

Thank you to Dr. Laurie Brogan and Dr. Holden Cox for their participation in our research study.

