Comparative Validation and Reliability of Artificial Intelligence in Countermovement Jump Assessment

Pablo Tadeo Ríos-Gallardo, PhD(c), MS.[1], Luis Enrique Carranza-García, PhD.[1], Carlos Balsalobre-Fernández, PhD.[2], Samuel Montalvo PhD., CPSS., CSCS, *D[3].

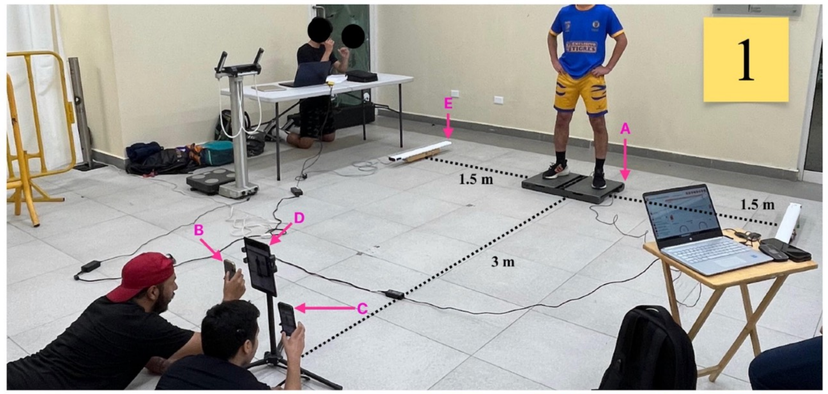
[1]School of Sport Organization, Autonomous University of Nuevo León, México., [2]Applied Biomechanics and Sports Technology Research Group, Autonomous University of Madrid, Spain., [3]Wu Tsai Performance Alliance, Division of Cardiovascular Medicine, Stanford University, California, USA.

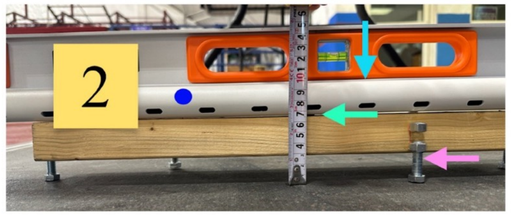
Background

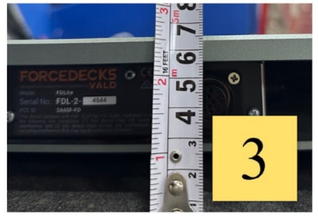
- The Countermovement Jump (CMJ) is widely used to assess athletic performance. While force platforms are the gold standard, their cost and lack of portability limit field use.
- Field methods like infrared systems and manual high-speed video (240 fps) estimate height via flight time but require precise setup and technique consistency.
- AI-based smartphone apps (e.g., 30 fps bounding-box tracking) offer accessible alternatives, yet lack robust validation against both laboratory and field methods across multiple athlete groups.

Objectives/Aims

- To assess the within- and between-session reliability, concurrent validity, and agreement of a 30 fps AI-based smartphone app (My Jump Lab AI) for CMJ height, compared simultaneously against a force platform, infrared timing, and 240 fps manual video across male athletes with varying performance levels.







Methods

1. Participants: Eighty healthy male university students (20.7 ± 1.7 yrs; 1.77 ± 0.07 m; 75.9 ± 11.7 kg) from four athletic groups (Recreational, Basketball, Soccer, Volleyball; *n*=20 each) completed two CMJ testing sessions, 7–10 days apart. Participants had no injuries and gave informed consent. The study was IRB-approved.

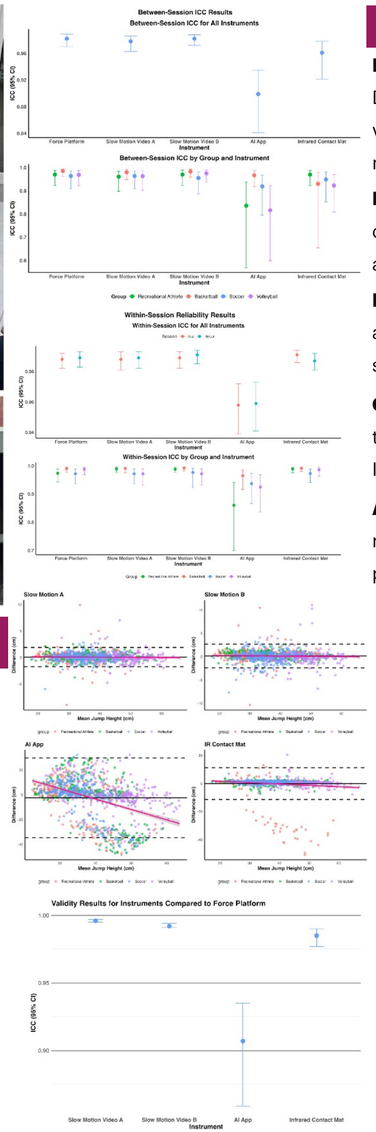
2. Protocol: Each session included a standardized warm-up, followed by three maximal CMJs (hands-on-hips). Two minutes of rest were provided between jumps. Participants wore the same attire and were tested at consistent times to control for variability.

3. Instrumentation: Five tools assessed CMJ height: **Force Platform (1000 Hz)** – impulse-momentum method. **Manual Video A/B (240 fps)** – flight-time via frame selection. **AI App (30 fps)** – bounding-box vertical displacement. **Infrared Mat (1000 Hz)** – flight-time via beam interruption.

4. Anthropometrics: Height, leg length, 90° squat height, and leverage were measured and input into My Jump Lab AI for displacement-to-height conversion. All assessments followed standardized measurement procedures.

5. Data Capture & Synchronization: All devices recorded each jump simultaneously under identical conditions. Manual videos were analyzed independently by two trained observers. The AI App included a built-in +2.5 cm correction for shoe thickness. Infrared beams were aligned precisely with the force platform surface.

6.Statistical Analysis: *Software:* RStudio (v. 2024.04.2), R (v. 4.4.2). **Reliability:** ICC(2,1), CV, SEM for within- and between-session consistency. **Validity:** ICC(2,1) vs. Force Platform. **Agreement:** Bland–Altman (mean bias, 95% LoA). **Proportional Bias:** Linear regression. **Packages:** *psych, irr, blandr, tidyverse, ggplot2.*



Results

Participant Characteristics: Eighty male athletes (n = 20/group) completed testing. Distinct group differences in jump height and physical profiles were observed, with volleyball players achieving the highest jumps and the AI App showing greater measurement variability, especially in recreational and volleyball groups.

Reliability *Within-Session:* Force Platform, Manual A/B, and Infrared Mat demonstrated excellent reliability (ICC ≥ 0.90; CV < 5%). The AI App showed acceptable ICCs but larger CVs, indicating greater variability and reduced precision.

Between-Session: Test-retest reliability was strong for Force Platform, Manual A/B, and Infrared Mat across all groups. The AI App exhibited lower consistency between sessions, with increased variation in group-specific performance.

Concurrent Validity: All alternative methods demonstrated strong agreement with the Force Platform. Manual A/B methods yielded the highest validity, followed by the Infrared Mat. The AI App showed the lowest, yet still acceptable, validity scores.

Agreement Analysis: Manual methods showed minimal bias and narrow agreement ranges. The AI App systematically underestimated jump height, with both fixed and proportional bias, lower prediction accuracy, and wider measurement dispersion.

Practical Applications

While AI-based apps offer accessible CMJ assessment, their higher variability and systematic bias warrant caution. Practitioners should avoid direct comparisons with kinetic or high-speed flight-time methods and select tools based on required precision, available resources, and the specific performance context.

Currículum Vitae

Figures & Supplemental Materials