

Efecto del nivel de experiencia clínica del examinador sobre la validez de criterio y fiabilidad inter-sesión de cinco medidas del rango de movimiento de la flexión dorsal del tobillo

Effect of the tester level of experience on the criterion-related validity and inter-session reliability of five different ankle dorsal flexion range of motion measures

Efeito da nível de experiência clínica do examinador sobre a validade de descrição e confiabilidade inter-sessão de cinco medidas da amplitude de movimento de la flexão dorsal do tornozelo

Fernando Gómez-Jiménez¹, Francisco Ayala², Antonio Cejudo¹, Pilar Sainz de Baranda³ y Fernando Santonja⁴

1 Master en Rendimiento Deportivo y Salud. Centro Deportivo INACUA, Murcia. 2 Doctor en Ciencias de la Actividad Física y del Deporte. Centro de Investigación del Deporte, Universidad Miguel Hernández de Elche. ISEN formación, centro adscrito a la Universidad de Murcia. 3 Doctora en Ciencias de la Actividad Física y del Deporte. Facultad del Deporte, Universidad de Murcia y 4 Doctor en Medicina y Cirugía. Facultad de Medicina. Universidad de Murcia.

Resumen: El objetivo de este estudio fue analizar el efecto que el diferente grado de experiencia del examinador tiene sobre la validez de criterio y fiabilidad inter-sesión de las medidas de rango de movimiento de la flexión dorsal del tobillo (ROM_{FDT}) obtenidas en las pruebas: flexión dorsal del tobillo pasiva en decúbito prono y rodilla flexionada (FD_{PAS}), zancada con inclinómetro ($FD_{ZAN-INC}$) y cinta métrica ($FD_{ZAN-MET}$), zancada modificada con taburete ($FD_{ZAN-TAB}$) y zancada modificada con cajón ($FD_{ZAN-CAJ}$). Un total de 17 deportistas recreativos completaron este estudio. Cada participante fue evaluado por dos examinadores con distinto nivel de experiencia (experto y novel) del ROM_{FDT} empleando 5 pruebas exploratorias diferentes en dos sesiones distintas. Los resultados muestran bajos valores de validez para el examinador novel en todas las pruebas de valoración del ROM_{FDT} excepto en la prueba $FD_{ZAN-INC}$ (error típico de la estimación estandarizado = 0,5 [pequeño]; Pearson = 0,85 [moderado]; kappa = 0,74 [buena]). El examinador experto presentó niveles altos de fiabilidad inter-sesión en las pruebas $FD_{ZAN-INC}$, $FD_{ZAN-MET}$ y $FD_{ZAN-CAJ}$ y valores aceptables en las pruebas FD_{PAS} y $FD_{ZAN-TAB}$. El examinador novel obtuvo valores aceptables de fiabilidad en todas las pruebas de valoración excepto en la prueba FD_{PAS} , que fueron considerados como no aceptables. Por lo tanto, se recomienda que tanto los examinadores expertos como noveles utilicen la prueba $FD_{ZAN-INC}$ para: (a) identificar a personas con restringida o normal movilidad de la flexión dorsal del tobillo; y (b) monitorizar la eficacia de los programas de intervención aplicados en caso de ser necesarios.

Palabras clave: Reproducibilidad, esguinces de tobillo, prevención de lesiones, precisión de la medida.

Abstract: The purpose of this study was to analyse the effect of the different degree of tester's experience on the criterion-related validity and the inter-session reliability of the ankle dorsiflexion range of motion measures (ROM_{ADF}) obtained from the following tests: passive non weight-bearing

test (ADF_{PAS}), weight-bearing lunge test using an inclinometer ($ADF_{LUN-INC}$) and a tape measure ($ADF_{LUN-TAP}$), modified-weight bearing test with a moveable datum ($ADF_{LUN-DAT}$) and modified weight bearing test with a box ($ADF_{LUN-BOX}$). A total of 17 recreational athletes completed this study. Each participant was assessed by 2 testers with different level of experience in the clinical setting (expert and novel) of the ROM_{ADF} using 5 assessment tests in two different sessions, with a rest-interval of 72-96 hours. The results showed low validity scores for the novel tester in all the assessment tests except for the $ADF_{LUN-INC}$ test (standardized typical error of the estimate = 0.5 [small] and Pearson coefficient = 0.85 [moderate]; Kappa = 0.74 [good]). The expert tester showed high reliability scores for the $ADF_{LUN-INC}$, $ADF_{LUN-TAP}$ and $ADF_{LUN-BOX}$ tests and acceptable scores for the ADF_{PAS} and $ADF_{LUN-DAT}$. The novel tester presented acceptable reliability scores for all the assessment tests except for the ADF_{PAS} , which were considered as non-acceptable. Therefore, we recommend that both the expert and novel testers use the $ADF_{LUN-INC}$ test to: (a) identify subjects with restrained or normal ROM_{ADF} and (b) monitor the efficacy of the intervention treatments applied in case of being necessary.

Key words: Reproducibility, ankle springs, injury prevention, accuracy of the measure.

Resumo: O objetivo deste estudo foi analisar o efeito de diferentes níveis de experiência do examinador tem sobre as medidas de validade de critério e inter-sessões variam de movimento da flexão dorsal do tornozelo (ROM_{FDT}) obtidos em testes de confiabilidade: flexão dorsal passiva do tornozelo propenso e joelho flexionado (FD_{PAS}), caminhar com inclinômetro ($FD_{ZAN-INC}$) e fita ($FD_{ZAN-MET}$), estribe fezes modificado ($FD_{ZAN-TAB}$) e estribe gaveta modificado ($FD_{ZAN-CAJ}$). Um total de 17 desportistas recreativos completaram o estudo. Cada participante foi avaliado por dois examinadores, com diferentes níveis de conhecimento (especialistas e novatos) de ROM_{FDT} utilizando 5 testes exploratórios diferentes em duas sessões diferentes. Os resultados mostram níveis baixos de validade para o romance examinador em todos os testes, exceto avaliação ROM_{FDT} do $FD_{ZAN-INC}$ (erro típico padrão da estimativa = 0,5 [pequena] test; Pearson = 0,85 [moderado]; kappa = 0,74 [boa]). O examinador especialista mostrou altos níveis de confiabi-

Dirección para correspondencia [Correspondence address]: Francisco Ayala. Centro de Investigación del Deporte, Universidad Miguel Hernández de Elche, Avda. de la Universidad s/n., 03202 Elche, Alicante (España). E-mail: Franciscoayalarodriguez@gmail.com

lidade inter-sessão de testes e valores aceitáveis em FD_{PAS} e testes $FD_{ZAN-TAB}$, $FD_{ZAN-INC}$, $FD_{ZAN-MET}$ e $FD_{ZAN-CAJ}$. Os valores de confiabilidade aceitável novela examinador obtidos em todos os testes, exceto teste de avaliação FD_{PAS} , as quais foram consideradas inaceitáveis. Portanto, recomenda-se que ambos os especialistas e novatos examinadores usar o teste $FD_{ZAN-INC}$:

(a) identificar as pessoas com mobilidade reduzida ou dorsiflexion normal de tornozelo; e (b) monitorar a eficácia de programas de intervenção ser aplicado, se necessário.

Palavras-chave: confiabilidade inter-sessão, entorses de tornozelo, prevenção de lesões, precisão da medição.

Introducción

La valoración del rango de movimiento de la flexión dorsal del tobillo (ROM_{FDT}) es una práctica habitual en el ámbito clínico y físico-deportivo porque un inadecuado nivel de movilidad ha sido propuesto como un factor importante que incrementa la probabilidad de sufrir numerosas patologías del tobillo y pie (DiGiovanni et al., 2002).

Existen un gran número de pruebas destinadas a la valoración del ROM_{FDT} , siendo las más utilizadas aquellas en las que el paciente o deportista es valorado de forma pasiva en decúbito prono con rodilla flexionada (FD_{PAS}) o de forma activa adoptando una posición de zancada hacia delante en bipedestación (FD_{ZAN}) (Krause, Cloud, Forster, Schrank, y Hollman, 2011; Martin y McPoil, 2005; Venturini, Ituassú, Teixeira y Deus, 2006). La prueba FD_{PAS} , más propia del ámbito clínico, requiere que el examinador a la vez que posiciona un inclinómetro o goniómetro sobre la cara lateral del 5º metatarsiano del tobillo del explorado realice un movimiento del tobillo hasta alcanzar el máximo ROM_{FDT} (figura 1) (Krause et al., 2011). Por su parte, la prueba FD_{ZAN} se basa en el principio de rodilla hacia la pared, y demanda que el explorado efectúe en frente de una pared un desplazamiento de la pelvis y tronco con la rodilla flexionada adoptando una posición de paso hacia delante (Bennell, Talbot, Wajswelner, Techovanich, Kelly y Hall, 1998). El objetivo de la prueba FD_{ZAN} es determinar el máximo ROM_{FDT} sin que el explorado levante el talón del suelo y con la cara anterior de la rodilla en contacto con la pared. Para determinar la máxima flexión dorsal del tobillo en la prueba FD_{ZAN} se han utilizado tanto el inclinómetro (figura 2; $FD_{ZAN-INC}$) como una cinta métrica pegada al suelo (figura 3; $FD_{ZAN-MET}$) (Bennell, et al., 1998; Denegar, Hertel y Fonseca, 2002; Grindstaff, Beazell, Magrum y Hertel, 2009; Hoch, Staton y McKeon, 2011; Vicenzino, Branjerdporn, Teys y Jordan, 2006). Además, recientemente se han propuesto algunas modificaciones de la prueba FD_{ZAN} , en donde se sustituye la pared por un taburete (figura 4; $FD_{ZAN-TAB}$) a empujar con la rodilla (O'Shea y Grafton, 2013) o un cajón (figura 5; $FD_{ZAN-CAJ}$) donde apoyar el pie (Cejudo, Sainz de Baranda, Ayala y Santonja, 2014a) para posteriormente registrar con un solo intento en ambas la máxima flexión dorsal del tobillo sin despegar el talón de la superficie donde se apoya.

En la actualidad no existe un consenso a la hora de establecer qué prueba exploratoria de las anteriormente citadas podría ser la más apropiada para valorar el ROM_{FDT} . Establecer una prueba exploratoria como la más apropiada o de referencia y fo-

mentar su uso en el ámbito clínico y físico-deportivo permitirá unificar criterios a la hora de valorar el ROM_{FDT} y diagnosticar posibles restricciones de movilidad. Además, en el ámbito científico permitirá ahondar con mayor celeridad en el estudio de posibles patologías del pie y tobillo asociadas a determinados valores de ROM_{FDT} (valores de normalidad y restricción) y analizar la eficacia de diferentes tratamientos destinados al mantenimiento y/o mejora de la movilidad del tobillo.

En este sentido, Hopkins (2000a y b, 2004) sugiere que la elección de pruebas diagnósticas de referencia debe de estar basada en una primera instancia en criterios de elevada validez y fiabilidad, para pasar después a valorar la sencillez y universalidad del procedimiento exploratorio.

En este sentido, las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ han sido aceptadas como válidas por las más prestigiosas Organizaciones Americanas de Medicina (American Academic of Orthopedic Association, 1965; American Medical Association, 2001) e incluidas en los manuales más acreditados del área de la Medicina y Ciencias del Deporte (Magee, 2002; Palmer y Epler, 2002; Prentice, 2003), basándose en los conocimientos anatómicos y en una amplia experiencia práctica. Con respecto a la fiabilidad de la medida de ROM_{FDT} obtenida en cada una de estas pruebas, estudios científicos informan de altos valores de fiabilidad inter-sesión (índice de correlación intraclase [ICC] = 0,93-0,98) (Bennell et al., 1998; Cejudo et al., 2014a y b; Konor, Morton, Eckerson y Grindstaff, 2012; O'Shea y Grafton, 2013; Simondson, Brock y Cotton, 2012; Venturini et al., 2006) para las pruebas $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ y moderados-altos (ICC = 0,70-0,95) (Krause et al., 2011) para la prueba FD_{PAS} .

Amparados en el hecho de que las pruebas $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ parecen presentar similares valores de validez y fiabilidad, determinados autores han expresado sus preferencias a la hora de valorar el ROM_{FDT} , basándose principalmente en criterios de sencillez del procedimiento exploratorio y universalidad de recursos materiales. Así, Cejudo et al. (2014a) sugieren el empleo de las pruebas $FD_{ZAN-CAJ}$ y $FD_{ZAN-INC}$ en lugar de las pruebas $FD_{ZAN-MET}$ y $FD_{ZAN-TAB}$ porque estas últimas implican la adopción por parte del examinador de posiciones muy incómodas para poder identificar el máximo grado de flexión dorsal del tobillo sin despegar el talón del suelo. Sin embargo, Bennell et al. (1998) defienden el empleo de las pruebas $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ porque únicamente requieren del empleo de una cinta métrica para ser llevadas a cabo.

Sin embargo, cuando se analiza en profundidad la litera-

tura científica se observa que la mayor parte de los estudios que han evaluado la validez y fiabilidad inter-sesión de las medidas obtenidas en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ utilizan examinadores con una gran experiencia en el ámbito clínico (Bennell et al., 1998; Cejudo et al., 2014a; Hoch y McKeon, 2011; Kaufman, Brodine, Shaffer, Johnson y Cullison, 1999; O'Shea y Grafton, 2013; Pope, Herbert y Kirwan, 1998; Simondson, Brock y Cotton, 2012), lo cual limita la utilidad de todas las recomendaciones anteriormente expuestas.

Un aspecto fundamental a la hora de determinar y categorizar el nivel de dificultad de un procedimiento exploratorio es analizar si examinadores noveles o con escasa experiencia en el ámbito clínico obtienen valores altos de correlación con respecto a la medida criterio (validez de criterio) establecida por parte de un examinador experimentado en la misma prueba (Hopkins, 2000a). Esta circunstancia evitará que examinadores noveles realicen diagnósticos inapropiados de normalidad o restringido ROM_{FDT} debido al empleo de pruebas demasiado complejas para ellos (Mayer Kondraske, Beals y Gatchel, 1997). Además, analizar y comparar la fiabilidad inter-sesión de las medidas de ROM_{FDT} obtenidas por examinadores noveles y experimentados en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$, $FD_{ZAN-CAJ}$ permitirá identificar cuáles de ellas presentan valores aceptables de estabilidad, además de valorar la influencia del sesgo de aprendizaje del examinador novel. Todo este conocimiento dotará a los profesionales de la Medicina y Ciencias del Deporte de argumentos evidenciados que permitan identificar qué prueba de valoración del ROM_{FDT} podría ser considerada como de referencia o criterio. Sorprendentemente, no existen (desde el conocimiento de los autores) estudios científicos que hayan analizado el impacto que el diferente nivel de experiencia o grado de entrenamiento del examinador presenta sobre la validez y fiabilidad inter-sesión de las medidas obtenidas en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$. Además, únicamente 2 estudios han examinado la fiabilidad inter-sesión de las medidas de ROM_{FDT} obtenidas en las pruebas $FD_{ZAN-INC}$, $FD_{ZAN-MET}$ por examinadores noveles, informando de valores moderados altos ($ICC = 0,85-0,96$) (Bennell et al., 1998; Konor et al., 2012). Por lo tanto, el objetivo principal de este estudio fue analizar el efecto que el diferente grado de experiencia del examinador (experto y novel) tiene sobre la validez de criterio y fiabilidad inter-sesión de las medidas obtenidas en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ en deportistas recreativos.

Método

Muestra

La muestra del presente estudio estuvo formada por dos examinadores con diferente grado de experiencia (experto y

novel) en el ámbito clínico. Así, el examinador considerado experto y cuya medida de ROM_{FDT} se estableció como criterio en las diferentes pruebas de valoración analizadas, fue un doctor en Ciencias de la Actividad Física y del Deporte con más de 10 años de experiencia en el ámbito clínico. El segundo examinador, considerado novel, fue un estudiante de Master, quien recibió únicamente un curso teórico-práctico de 2 horas de duración sobre las diferentes pruebas e instrumentos de valoración del ROM_{FDT} .

Procedimiento

Un total de 17 hombres adultos jóvenes (edad: $21,4 \pm 2,9$ años; estatura: $175,3 \pm 3,9$ cm; masa corporal: $70,2 \pm 6,7$ kg) físicamente activos (3-5 sesiones de actividad física semanales de al menos 1 hora de duración) formaron parte en el presente estudio como participantes que debían ser evaluados por ambos examinadores. Aunque todos los participantes eran deportistas recreativos, ninguno de ellos realizaba un entrenamiento sistemático y específico de fuerza y flexibilidad de la extremidad inferior. Todos los participantes fueron instados a mantener su régimen habitual de práctica deportiva durante toda la fase experimental del estudio, evitando actividades físicas de intensidad vigorosa las 48 horas previas a cada sesión de valoración.

Los criterios de exclusión fueron: (a) presentar alteraciones músculo-esqueléticas, tales como desgarros del tríceps sural y tendón de Aquiles en los últimos 6 meses previos al presente procedimiento exploratorio; (b) presentar dolor muscular de aparición tardía (agujetas) durante cualquiera de los dos momentos de evaluación; y (c) no asistir a una o más sesiones de valoración durante todo el proceso de recogida de datos.

Todos los participantes fueron verbalmente informados de la metodología a utilizar, así como de los propósitos y posibles riesgos del estudio, y un consentimiento informado fue firmado por cada uno de ellos. El presente estudio fue aprobado por el Comité Ético y Científico de una Universidad.

Todos los participantes fueron evaluados por cada uno de los dos clínicos del ROM_{FDT} en dos sesiones diferentes, con un periodo de tiempo entre sesiones de 5-7 días. En cada una de las sesiones, cada participante fue evaluado aleatoriamente por el clínico novel y experto, quienes estuvieron situados en salas diferentes, bajo las mismas condiciones ambientales.

Cada clínico evaluó de forma aleatoria el ROM_{FDT} a través de las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$. Únicamente se evaluó el ROM_{FDT} de la pierna dominante, ya que estudios previos no han observado asimetrías bilaterales en muestras de adultos jóvenes físicamente activos y asintomáticos (Stefanyshyn y Engsborg, 1994; Scharfbillig y Scutter 2004). Cada participante realizó dos intentos máximos para cada una de las pruebas de valoración, seleccionándose la media para el posterior análisis estadístico. Se realizó

un descanso de aproximadamente 30 segundos entre cada uno de los dos intentos máximos para cada prueba, de 60 segundos entre pruebas de valoración y de 10 minutos entre examinadores. Todos los participantes fueron evaluados en ropa deportiva y descalzos. Dos investigadores independientes y ajenos al objetivo del estudio fueron los encargados de registrar las medidas de ROM_{FDT} obtenidas por los dos examinadores en cada una de las pruebas de valoración y sesiones.

Figura 1. Test de flexión dorsal del tobillo en decúbito prono con rodilla flexionada (FD_{PAS}).



Figura 2. Test de la zancada original en centímetros ($FD_{ZAN-MET}$).



Figura 3. Test de la zancada original en grados ($FD_{ZAN-INC}$).

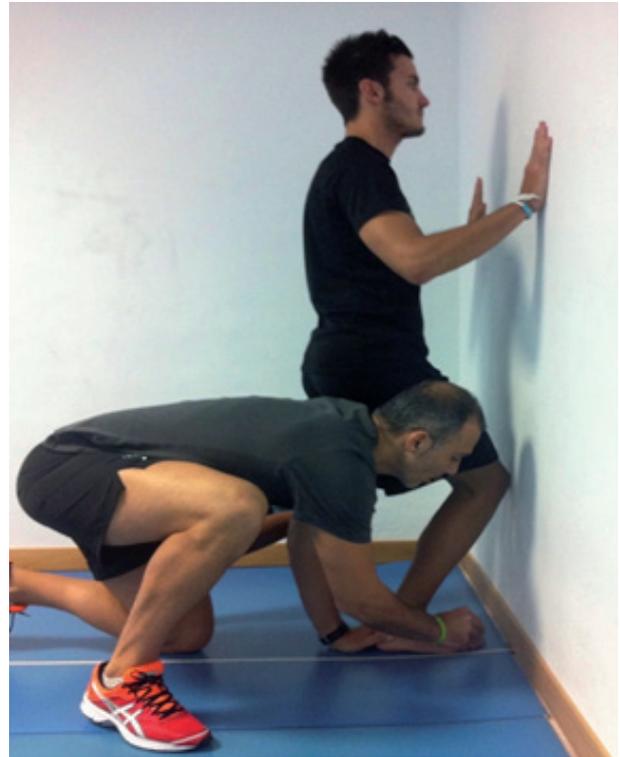


Figura 4. Modificación del test de la zancada propuesta por O'Shea et al. (2013) ($FD_{ZAN-TAB}$).



Figura 5. Modificación del test de la zancada propuesta por Cejudo et al. (2014a) ($FD_{ZAN-CAJ}$).



Medidas

Las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ fueron llevadas a cabo siguiendo el procedimiento exploratorio descrito por Bennell et al. (1998), Konor et al. (2012), O'Shea y Grafton (2013) y Cejudo et al. (2014a y b) respectivamente. Un inclinómetro ISOMED con una varilla telescópica adherida a uno de sus extremos fue utilizado para medir el ROM_{FDT} en las pruebas FD_{PAS} , $FD_{ZAN-INC}$ y $FD_{ZAN-CAJ}$. El inclinómetro fue calibrado con la vertical antes y durante cada una de las sesiones de valoración para asegurar que no se produjesen cambios en su sensibilidad. Una cinta métrica adherida al suelo fue utilizada para medir en las pruebas $FD_{ZAN-MET}$ y $FD_{ZAN-TAB}$ la máxima distancia en centímetros existente entre la falange más distal del pie y la pared sin despegar el talón del suelo y con la rodilla en contacto con la pared o taburete respectivamente.

Para un mejor entendimiento del método de evaluación de cada una de las pruebas de valoración del ROM_{FDT} seleccionadas (posición inicial, calibración del inclinómetro, movimiento, posición final y medición) se presenta una descripción visual de cada una en las figuras 1-5.

Análisis estadístico

El análisis de los datos se llevó a cabo a través del software estadístico SPSS versión 20.0 para Windows. La distribución

normal de los datos fue comprobada a través de la prueba Kolmogorov-Smirnov, demostrando todos ellos una distribución normal ($p > 0,05$). Se realizó un análisis descriptivo de cada una de las variables, que incluía la media y su correspondiente desviación típica. Un análisis de la varianza 2 (examinador novel y examinador experto) \times 5 (FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$) \times 2 (sesión 1 y sesión 2) se llevó a cabo para determinar la existencia de diferencias estadísticamente significativas en los resultados obtenidos por los 2 examinadores en cada una de las diferentes pruebas y sesiones de valoración. En el caso de identificar diferencias significativas, se aplicó la prueba post hoc Bonferroni para comparaciones pareadas. El nivel de significación se estableció en $p < 0,05$.

Validez

La validez de criterio de las medidas de ROM_{FDT} obtenidas por el examinador novel en la primera sesión de valoración fue determinada a través de una ecuación de estimación, el error típico de la estimación (ET_{EST}) y un análisis de correlación (coeficiente de Pearson), empleando para ello la metodología descrita por Hopkins (2000a). Así, la ecuación de estimación fue calculada como la ecuación generada tras representar gráficamente y después establecer una línea recta los valores de ROM_{FDT} del examinador experto con los del examinador novel para cada una de las pruebas de valoración ($y =$ pendiente $\times X +$ intersección). El ET_{EST} fue calculado como la media del error típico de la diferencia entre los valores obtenidos por ambos examinadores en cada una de las pruebas para cada uno de los participantes, y fue expresado como desviación estándar. Para interpretar los valores de ET_{EST} , Hopkins (2000a) sugiere calcular el ET_{EST} estandarizado (ET_{EST} / SD de la medida criterio ([examinador experimentado]) y posteriormente aplicar la siguiente escala de valores: $< 0,2$ trivial; $0,2-0,6$ pequeño; $0,6-1,2$ moderado; $1,2-2,0$ largo y $> 2,0$ muy largo. La validez fue también estudiada a través del coeficiente de correlación de Pearson (β) entre las medidas obtenidas por ambos clínicos en cada una de las pruebas de valoración. El grado de correlación fue determinado utilizando la siguiente escala de valores (Hopkins, 2000a): $< 0,80$ bajo, $0,80-0,90$ moderado y $> 0,90$ alto. El estudio de la posible existencia de heterocedasticidad de los datos en cada una de las pruebas de valoración se llevó a cabo mediante el análisis gráfico y de correlación del grado de dispersión de los valores residuales y predictivos (Hopkins, 2000a). El análisis de la magnitud del sesgo error ("systematic bias") entre los valores predictivos (examinador novel) con respecto a los valores criterio (examinador experto) para cada una de las pruebas de valoración se llevó a cabo a través del cálculo de su diferencia y desviación típica de la diferencia así como su posterior representación gráfica siguiendo la estructura propuesta por Bland & Altman (1986). Para la interpretación de la magnitud del sesgo error de cada una de las pruebas se llevó a cabo su

transformación en valores estandarizados (sesgo error estandarizado = sesgo error en valores absolutos / desviación estándar de la media de los valores totales registrados por el examinador experto) y su posterior traducción cualitativa siguiendo la escala propuesta por Hopkins (2000a): <0,2 trivial; 0,2-0,6 pequeño; >0,6-1,2 moderado; >1,2-2,0 grande; >2,0 muy grande.

Finalmente, los valores obtenidos por cada uno de los examinadores en cada una de las pruebas y para cada uno de los participantes fueron dicotomizados para indicar un valor positivo o negativo de acuerdo a los valores de corte de normalidad o restringido ROM_{FDT} propuestos por Kaufman et al. (1999) para la prueba FD_{PAS} (<18,5° = restringido), Pope, Herbert y Kirwan (1998) para las medidas angulares de la prueba FD_{ZAN-INC} y FD_{ZAN-CAJ} (<34° = restringido) y Bennell et al. (1998) para las medidas lineales de la prueba FD_{ZAN-MET} y FD_{ZAN-TAB} (<10 cm = restringido). Tras reducir los resultados a variables nominales (positivo = restringido ROM_{FDT}; negativo = normal ROM_{FDT}), el índice de correlación de kappa fue calculado para determinar el grado de acuerdo entre los dos examinadores (novel y experto) para clasificar a los participantes como personas con un restringido o normal ROM_{FDT}. Los resultados obtenidos en la correlación de kappa fueron interpretados de acuerdo a la siguiente escala: <0,20 pobre; 0,21-0,40 débil; 0,41-0,60 moderada; 0,61-0,80 buena y 0,81-1,00 muy buena (Landis y Koch, 1977).

Fiabilidad inter-sesión

La fiabilidad relativa de cada una de las medidas de ROM_{FDT} obtenidas a través de las diferentes pruebas de valoración se determinó a través del Índice de Correlación Intraclase modelo 2k (ICC_{2,k}). La elección del modelo 2k se justifica en el hecho de que permite determinar si una medida puede ser utilizada con seguridad y fiabilidad entre clínicos con característica similares a los seleccionados en el presente estudio (Weir, 2005). La interpretación cualitativa de los valores del estadístico ICC_{2k} se basó en la propuesta establecida por Hopkins (2000b), quién considera valores bajos los <0,80; moderados los que oscilan entre 0,80-0,90 y altos los mayores de 0,90.

La fiabilidad absoluta de las medidas de la flexión dorsal

del tobillo se determinó a través de los estadísticos cambio en la medida (CM), estándar error de la medida (SEM) y mínimo cambio detectable con un intervalo de confianza del 95% (MDC₉₅) (Atkinson y Nevill, 1998; Weir, 2005).

El CM de las diferentes medidas de flexión dorsal del tobillo entre sesiones de evaluación y examinadores fue determinado empleando la metodología y hoja de cálculo desarrollada por Hopkins (2007) a través del estadístico t para varianzas iguales. Para la realización de inferencia sobre el valor real del efecto, la incertidumbre del efecto se expuso a través de su intervalo de confianza al 90% y como probabilidad de que el verdadero valor del efecto represente un cambio sustancial (negativo o positivo). La probabilidad de que el CM fuese positivo o negativo se interpretó de acuerdo a la siguiente escala: <0,5%, seguramente trivial; 1-5%, muy probablemente trivial; 6-25%, probablemente trivial; 26-75%, posiblemente positivo o negativo; 76-95%, probablemente positivo o negativo; 95-99%, muy probablemente positivo o negativo; > 99% seguramente positivo o negativo (Hopkins, Marshall, Batterham y Hanin, 2009).

El estadístico SEM se calculó empleando los valores “crudos” de las medidas obtenidas en las pruebas FD_{PAS}, FD_{ZAN-INC}, FD_{ZAN-MET}, FD_{ZAN-TAB} y FD_{ZAN-CAJ} a través de la siguiente fórmula: $\sqrt{\text{MSE}}$, donde MSE es la media cuadrática del error derivada de análisis de la varianza con medidas repetidas (Weir, 2005). El MDC₉₅ fue calculado como $\text{SEM} \times \sqrt{2} \times 1.96$ (Weir, 2005). Tanto el SEM como el MDC₉₅ fueron expresados tanto en valores absolutos como porcentuales con respecto a la media de la medida de la primera sesión de valoración (%SEM y %MDC₉₅).

Resultados

En la tabla 1 se presenta la estadística descriptiva (media \pm desviación estándar) de las diferentes pruebas de valoración del ROM_{FDT} separadas según sesión de valoración y examinador (novel y experto). El análisis de la varianza mostró la no existencia de diferencias estadísticamente significativas ($p > 0,05$) entre los valores obtenidos por ambos examinadores para cada una de las pruebas de valoración del ROM_{FDT} y sesiones de evaluación.

Tabla 1. Estadística descriptiva (media \pm desviación estándar) de las diferentes pruebas de valoración separadas según sesión y clínico.

Prueba de Valoración	Sesión de valoración 1		Sesión de valoración 2	
	Clínico Experto	Clínico Novel	Clínico Experto	Clínico Novel
FD _{ZAN-MET} (cm)	24,7 \pm 5,6	26,8 \pm 7	25,2 \pm 4,4	30,6 \pm 6,3
FD _{ZAN-INC} (°)	9,2 \pm 3,2	9,6 \pm 2,5	8,9 \pm 3,2	9,2 \pm 2,6
FD _{ZAN-TAB} (cm)	39,7 \pm 4,5	40,7 \pm 5,2	38 \pm 4,7	38,4 \pm 5,4
FD _{ZAN-CAJ} (°)	8,8 \pm 3,4	9,6 \pm 2,6	8,7 \pm 3,3	9,5 \pm 2,3
FD _{PAS} (°)	36,5 \pm 5,2	39,2 \pm 5,7	35,2 \pm 5,4	36,6 \pm 5,7

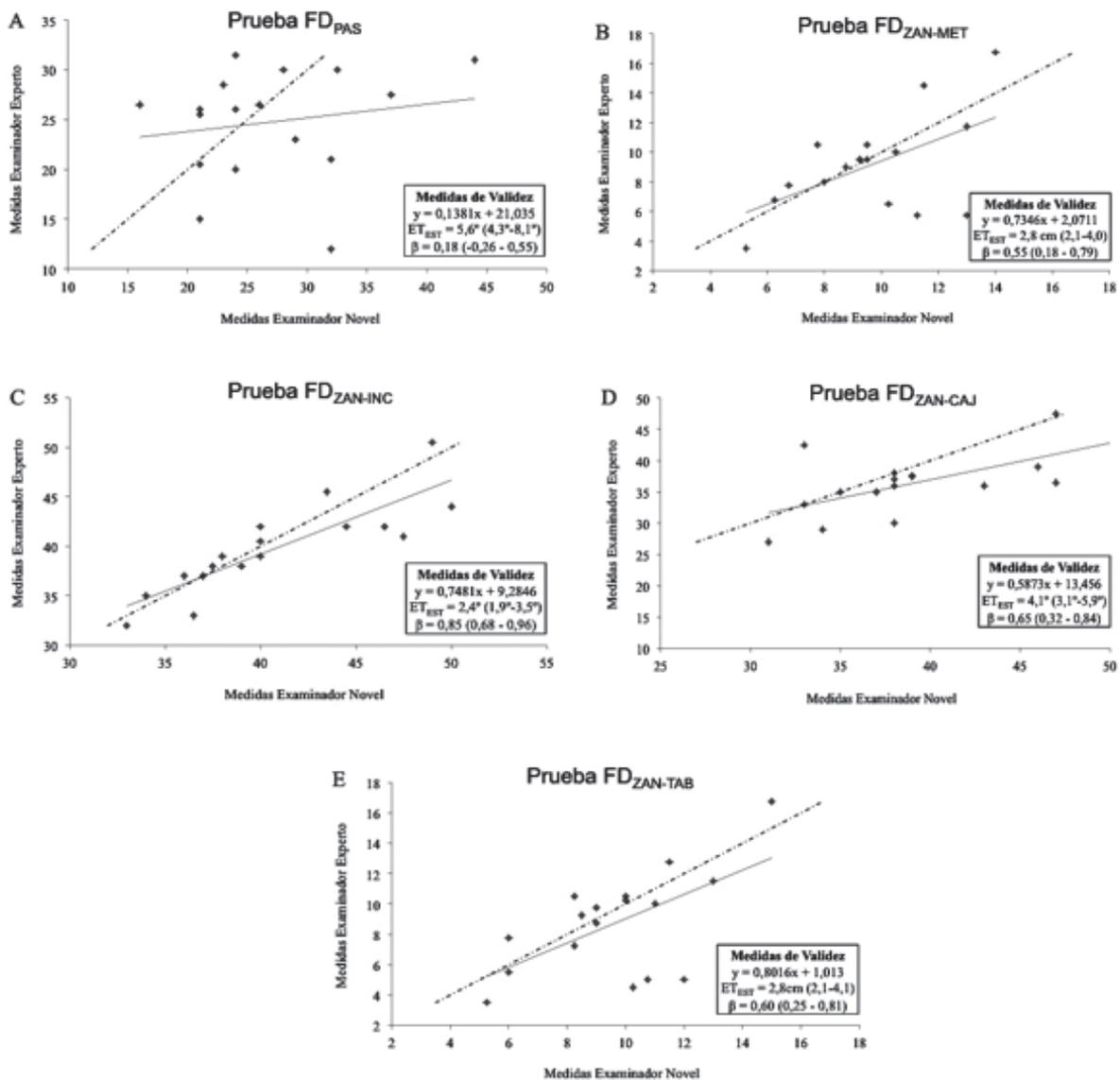
FD_{ZAN-MET}: test de la zancada original en centímetros; FD_{ZAN-INC}: test de la zancada original en grados; FD_{ZAN-TAB}: modificación del test de la zancada propuesta por O'Shea et al.¹⁰; FD_{ZAN-CAJ}: modificación del test de la zancada propuesta por Cejudo et al.¹¹; FD_{PAS}: test de flexión dorsal del tobillo en decúbito prono con rodilla flexionada; °: grados; cm: centímetros.

Validez

Los estadísticos de validez de cada una de las pruebas de valoración se presentan en la figura 6 (ecuación de estimación, ET_{EST} y β) y tabla 2 (correlaciones de Kappa). Pobre valores

de validez de criterio fueron encontrados para el examinador novel en todas las pruebas de valoración (ET_{EST} estandarizado $> 0,8$ [moderado-largo]; $\beta < 0,65$ [bajos]) excepto en la prueba $FD_{ZAN-INC}$ (ET_{EST} estandarizado = $0,5$ [pequeño] y $\beta = 0,85$ [moderado]).

Figura 6. Los valores de rango de movimiento de la flexión dorsal del tobillo obtenidos en las diferentes pruebas de valoración por el examinador experto fueron representados gráficamente con los obtenidos por el examinador novel para cada uno de los participantes. La línea discontinua es la línea de identidad (validez perfecta), y la línea sólida es la línea recta de los valores de la prueba FD_{PAS} (gráfico A), $FD_{ZAN-MET}$ (gráfico B) $FD_{ZAN-INC}$ (gráfico C), $FD_{ZAN-CAJ}$ (gráfico D) o $FD_{ZAN-TAB}$ (gráfico E). Cualquier desviación de la línea discontinua representa un desplazamiento sistemático. Las validez de criterio está representada en el córner inferior derecho de cada uno de los gráficos, empleando para ello los estadísticos: ecuación de estimación ($y =$ pendiente $\times X +$ intersección), ET_{EST} y el coeficiente de Pearson (β), y su correspondiente 90% intervalo de confianza.



Además, solo los resultados dicotomizados obtenidos en la prueba $FD_{ZAN-INC}$ por ambos examinadores obtuvieron un nivel de acuerdo aceptable desde el punto de vista clínico ($k = 0,74$ [buena]). El análisis gráfico del grado de dispersión de los valores residuales y predictivos mostró la no existencia de heterostedasticidad de los datos para cada una de las pruebas ($r < 0,1$; $p > 0,6$).

Finalmente, el análisis de la magnitud del sesgo error estandarizado de los valores del examinador novel con respecto a los valores del examinador criterio no fue en ningún caso superior a 0,6 (pequeño) (figura 7).

Tabla 2. Correlaciones de Kappa.

	FD_{PAS}	$FD_{ZAN-MET}$	$FD_{ZAN-INC}$	$FD_{ZAN-TAB}$	$FD_{ZAN-CAJ}$
FD_{PAS}	-0,1				
$FD_{ZAN-MET}$		0,38			
$FD_{ZAN-INC}$			0,74		
$FD_{ZAN-TAB}$				0,28	
$FD_{ZAN-CAJ}$					0,46

$FD_{ZAN-MET}$: test de la zancada original en centímetros; $FD_{ZAN-INC}$: test de la zancada original en grados; $FD_{ZAN-TAB}$: modificación del test de la zancada propuesta por O’Shea et al.¹⁰; $FD_{ZAN-CAJ}$: modificación del test de la zancada propuesta por Cejudo et al.¹¹; FD_{PAS} : test de flexión dorsal del tobillo en decúbito prono con rodilla flexionada.

Figura 7. Análisis de la magnitud del sesgo error entre los valores obtenidos por el examinador novel y experto para cada una de las pruebas. Se presentan los gráficos Bland-Altman para cada una de las pruebas de valoración del ROM_{FDT} (FD_{PAS} gráfico A; $FD_{ZAN-MET}$ gráfico B; $FD_{ZAN-INC}$ gráfico C; $FD_{ZAN-CAJ}$ gráfico D; $FD_{ZAN-TAB}$ gráfico E). Los valores pertenecen a la media de los resultados obtenidos en las dos sesiones de valoración efectuadas. El análisis de la magnitud del sesgo error estandarizado (SE_{EST}) de los valores predictivos (examinador novel) con respecto a los valores criterio (examinador experto) muestran valores inferiores a 0,6 (pequeño).

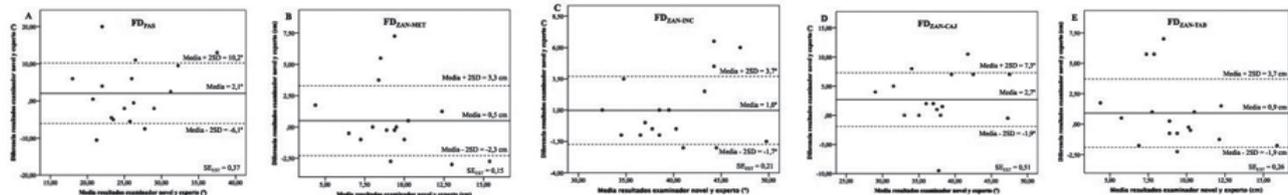


Tabla 3. Estadísticos de fiabilidad inter-sesión para cada uno de las pruebas de valoración y examinador (experto y novel): cambio en la media (CM), error estándar de la medida (SEM), cambio mínimo detectable con un nivel de confianza del 95% (MDC_{95}) e índice de correlación intraclass (ICC_{2k}).

Test	Examinador Experto						Examinador Novel					
	CM	SEM	%SEM	MDC_{95}	% MDC_{95}	ICC	CM	SEM	%SEM	MDC_{95}	% MDC_{95}	ICC
$FD_{ZAN-MET}$ (cm)	-0,5*	1,9	4,5	5,3	12,4	0,98	-0,5*	0,6	5,8	1,6	16,3	0,96
$FD_{ZAN-INC}$ (°)	-0,7*	1	2,4	2,6	6,7	0,91	-2,3*	2,2	5,4	6,1	15,1	0,91
$FD_{ZAN-TAB}$ (cm)	0,1*	0,5	6,1	1,5	16,8	0,98	-0,1*	0,7	7,5	2	20,8	0,85
$FD_{ZAN-CAJ}$ (°)	-1,2*	1,4	3,8	3,9	10,6	0,93	-1,8§	2,5	6,5	7,1	18,1	0,85
FD_{PAS} (°)	0,5*	1,9	7,5	5,2	20,1	0,92	3,8§	3,3	12,3	9,1	34,3	0,72

$FD_{ZAN-MET}$: test de la zancada original en centímetros; $FD_{ZAN-INC}$: test de la zancada original en grados; $FD_{ZAN-TAB}$: modificación del test de la zancada propuesta por O’Shea et al.¹⁰; $FD_{ZAN-CAJ}$: modificación del test de la zancada propuesta por Cejudo et al.¹¹; FD_{PAS} : test de flexión dorsal del tobillo en decúbito prono con rodilla flexionada; °: grados; cm: centímetros; *: posiblemente trivial; §: posiblemente negativo o positivo.

Fiabilidad inter-sesión

Los estadísticos de fiabilidad (CM, SEM, %SEM, %MDC e ICC) de las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ se muestran en la tabla 3 de forma separada para el examinador experto y novel.

Para el examinador experto, el estadístico CM muestra que las diferencias en los valores de ROM_{FDT} entre las dos sesiones de valoración son posiblemente triviales para todas las pruebas de valoración. Los estadísticos SEM y MDC_{95} de cada media de ROM_{FDT} oscilaron entre 0,5 y 1,9 y entre 1,5 y 5,3 respectivamente, con valores de %SEM inferiores al 5% (excepto para las pruebas FD_{PAS} y $FD_{ZAN-TAB}$) y con valores de ICC mayores de 0,90 (altos). Para el examinador novel, se observan diferencias posiblemente positivas o negativas en los valores de ROM_{FDT} entre sesiones de valoración para las pruebas FD_{PAS} y $FD_{ZAN-CAJ}$.

Los estadísticos SEM y MDC_{95} de cada medida de ROM_{FDT} oscilaron entre 0,6 y 3,3 y entre 1,6 y 9,1 respectivamente, con valores del %SEM entre 5,4 y 12,3 y con valores de ICC entre 0,72 (bajos) y 0,96 (altos).

Discusión

Las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ son las más utilizadas en el ámbito clínico y físico-deportivo para valorar el ROM_{FDT} . Sin embargo, y dada la disparidad de sus procedimientos exploratorios, es posible que la aplicación de algunas de estas pruebas pueda estar supeditada a examinadores con un alto nivel de experiencia o entrenamiento. Así, el empleo de pruebas de valoración del ROM_{FDT} demasiado complejas por parte de examinadores noveles o con poco dominio técnico podría derivar en falsos diagnósticos de movilidad restringida o normal y/o en inapropiadas acciones de monitorización o seguimiento. Es por ello que el objetivo principal de este estudio fue analizar el impacto que el diferente grado de entrenamiento (experto y novel) presenta sobre la validez de criterio y fiabilidad inter-sesión de las medidas de ROM_{FDT} obtenidas en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ en deportistas recreativos.

Con respecto a la validez de criterio de las medidas de ROM_{FDT} analizadas, los hallazgos de este estudio sugieren que únicamente la prueba $FD_{ZAN-INC}$ mostró valores moderados de validez de acuerdo a la categorización establecida en la literatura científica para los estadísticos ET_{EST} estandarizado (0,5 [pequeño]) y coeficiente de Pearson (0,85 [moderado]) (Hopkins, 2000b). Además, el análisis de correlación de Kappa mostró que la prueba $FD_{ZAN-INC}$ fue la única que obtuvo valores de acuerdo aceptables entre el examinador novel y experto para categorizar a un sujeto con un ROM_{FDT} restringido o normal ($k = 0,74$ [buena]). Sin embargo, antes de

considerar como válida la medida de ROM_{FDT} obtenida en la prueba $FD_{ZAN-INC}$ por el examinador novel, es importante decidir si el umbral del valor predictivo conseguido tras aplicar en la ecuación de estimación un valor aleatorio de ROM_{FDT} obtenido por el examinador novel es lo suficientemente pequeño para detectar con precisión el diagnóstico deseado (e.j. ROM_{FDT} restringido). Hopkins (2000b) sugiere que un rango representado por $\pm 1,0$ (68% probabilidad o probabilidad 2-1 de que contenga el verdadero valor) o $\pm 1,28$ (80% probabilidad o probabilidad 4-1 de que contenga el verdadero valor) $\times ET_{EST}$ podría ser apropiado para decidir si una medida es lo suficientemente válida y precisa para diagnosticar. Así, cuando se calculó el umbral del valor predictivo para la medida de ROM_{FDT} obtenida en la prueba $FD_{ZAN-INC}$ por el examinador novel utilizando como valor aleatorio 34° (punto de corte de movilidad restringida) y el enfoque más restringido (1,28), se obtuvieron unos rangos que se podrían considerar aceptables desde el punto de vista clínico ($34,6 \pm 3^\circ$).

Una posible explicación al hecho de que únicamente la medida de ROM_{FDT} obtenida en la prueba $FD_{ZAN-INC}$ por el examinador novel obtuviese valores de validez aceptables desde el punto de vista clínico puede radicar en que presenta el procedimiento exploratorio más sencillo. En este sentido, al tener como referencia la pared, el examinador únicamente tiene que prestar atención a la correcta colocación del pie que esta siendo evaluado, a diferencia de la prueba $FD_{ZAN-TAB}$, donde hay que monitorizar además el movimiento del taburete. Igualmente, el empleo del inclinómetro como instrumento de medida evita que el examinador tenga que adoptar posiciones demasiado incómodas a la hora de registrar la máxima flexión dorsal del tobillo, a diferencia de la prueba $FD_{ZAN-MET}$, donde se adoptan posiciones de máxima flexión de tronco, que podrían impedir la correcta visualización de la distancia en centímetros por adoptar planos laterales y no completamente cenitales. Finalmente, la utilización de la bisectriz de la pierna como referencia de donde colocar la varilla telescópica evita que el examinador tenga que identificar marcas corporales específicas como la tuberosidad de la tibia (Munteanu, Strawhorn, Landorf, Bird y Murley, 2009) o la zona más distal de la tibia (Denegar, Hertel y Fonseca, 2002), favoreciendo así la consistencia entre examinadores.

Por otro lado, los hallazgos de este estudio también demuestran que un examinador novel o con escaso entrenamiento presenta peores valores de fiabilidad inter-sesión en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ comparado con un examinador experto. Determinador autores sugieren que un "apropiado" nivel de fiabilidad inter-sesión podría ser definido cuando una medida presenta una alta estabilidad (CM no significativo [ausencia de error sistemático]), una variabilidad (SEM) inferior al 5-10% y un ICC $> 0,80-0,90$ (Atkinson, Nevill, y Edwards, 1999, Hopkin, 2000a). En este sentido, y desde la perspectiva más ambiciosa,

los resultados de este estudio muestran como el examinador experto presentó niveles altos de fiabilidad inter-sesión ($SEM < 5\%$ e $ICC > 0,9$) en las medidas de ROM_{FDT} obtenidas en las pruebas $FD_{ZAN-INC}$, $FD_{ZAN-MET}$ y $FD_{ZAN-CAJ}$, mientras que valores aceptables ($SEM = 5-10\%$; $ICC = 0,8-0,9$) fueron encontrados en las pruebas FD_{PAS} y $FD_{ZAN-TAB}$. Por su parte, el examinador novel únicamente obtuvo valores aceptables de fiabilidad inter-sesión en las medidas de ROM_{FDT} obtenidas en las pruebas $FD_{ZAN-MET}$, $FD_{ZAN-INC}$, $FD_{ZAN-CAJ}$ y $FD_{ZAN-TAB}$ y no aceptables en la prueba FD_{PAS} . Las medidas de ROM_{FDT} obtenidas en las pruebas FD_{PAS} y $FD_{ZAN-TAB}$ fueron las que presentaron valores más bajos de fiabilidad inter-sesión tanto en el examinadores experto ($\%SEM = 6,1 [FD_{ZAN-TAB}] - 7,5 [FD_{PAS}]$; $ICC = 0,92 [FD_{PAS}] - 0,93 [FD_{ZAN-TAB}]$) como en el novel ($\%SEM = 7,5 [FD_{ZAN-TAB}] - 12,3 [FD_{PAS}]$; $ICC = 0,72 [FD_{PAS}] - 0,85 [FD_{ZAN-TAB}]$).

Quizás una posible explicación al hecho de que las pruebas FD_{PAS} y $FD_{ZAN-TAB}$ presenten valores de fiabilidad inter-sesión más bajos en ambos examinadores, experto y novel, pueda radicar en la mayor dificultad de sus procedimientos exploratorios. En este sentido, en la prueba FD_{PAS} el examinador necesita de una gran habilidad que le permita ejercer con una mano siempre la misma fuerza a la hora de movilizar la articulación del tobillo hacia la máxima flexión dorsal (Bennell et al., 1998) mientras que con la otra mano debe de colocar el inclinómetro sobre la bisectriz del borde externo del pie (figura 1). Así, cualquier pequeña variación en esta capacidad de ejercer siempre la misma fuerza (ej.: fatiga, nivel de motivación) entre sesiones de valoración podría alterar la fiabilidad de la medida. Además, en la prueba $FD_{ZAN-TAB}$ el hecho de tener que mover o desplazar un taburete con la parte anterior de la rodilla hasta alcanzar la máxima flexión dorsal del tobillo podría producir un movimiento de pronación del tobillo por parte del examinado para facilitar el movimiento del taburete y alcanzar una mayor distancia (Bohannon, Tiberio y Waters, 1991; Tiberio, Bohannon y Zito, 1989), lo que podría alterar la fiabilidad de la medida sino se monitoriza correctamente entre sesiones de valoración. Además, Cejudo et al. (2004a) consideran que la prueba $FD_{ZAN-TAB}$ podría producir en determinados pacientes o deportistas dolor en la cara anterior de la rodilla al tener que ejercer presión sobre el taburete para su movilización, lo cual podría cuestionar su utilidad. Todos estas dificultades metodológicas podrían haber afectado en mayor magnitud a la medida de ROM_{FDT} obtenida por el examinador novel, explicando así sus peores resultados de fiabilidad inter-sesión en las pruebas FD_{PAS} y $FD_{ZAN-TAB}$.

Similares resultados de fiabilidad inter-sesión han sido encontrados en estudios previos para las medidas de ROM_{FDT} obtenidas por un examinador experto en las pruebas FD_{PAS} (Beazell, Grindstaff, Sauer, Magrum, Ingersoll y Hertel, 2012; Grindstaff et al., 2009), FD_{ZAN} (Bennell et al., 1998; Konor et al., 2012), $FD_{ZAN-TAB}$ (O'Shea y Grafton, 2013) y $FD_{ZAN-CAJ}$ (Cejudo et al., 2014a y b). Por su parte, los valores de fiabilidad inter-sesión observados en el examinador novel son también similares a los encontrados por estudios previos para las pruebas $FD_{ZAN-INC}$ y $FD_{ZAN-MET}$ (Konor et al., 2012; Bennell et al., 1998). Desafortunadamente, no es posible comparar los resultados de fiabilidad inter-sesión obtenidos por el examinador novel en las pruebas FD_{PAS} , $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$ con estudios previos, ya que el presente estudio es el primero que ha abordado esta cuestión.

Aunque el presente estudio es el primero en abordar el impacto del grado de entrenamiento del examinador sobre la fiabilidad y validez de la medida de ROM_{FDT} obtenida en diferentes pruebas de valoración, ciertas limitaciones deben ser aclaradas. En primer lugar, el escaso tamaño muestral ($n = 2$) y el limitado rango de edad de los participantes podría limitar la validez de los resultados obtenidos. Además, si estos resultados son extrapolables a personas con patologías de tobillo o pie y distinto nivel de entrenamiento al empleado en este estudio es un aspecto que se escapa del alcance de este estudio. Como futuras líneas de investigación se propone llevar a cabo estudios que analicen el efecto del nivel de experiencia clínica sobre la precisión de las medidas de ROM_{FDT} empleando varios examinadores expertos y examinadores noveles con diferente nivel de experiencia (ej.: novel con escasa formación práctica vs novel con una formación práctica de 20 horas,...) para establecer la evolución del aprendizaje de estos últimos y establecer así recomendaciones sobre el diseño de programas de formación.

Aplicaciones prácticas

Los hallazgos de este estudio advierten de que el nivel de experiencia del examinador podría influir en la precisión de las medidas de ROM_{FDT} obtenidas en las pruebas FD_{PAS} , $FD_{ZAN-INC}$, $FD_{ZAN-MET}$, $FD_{ZAN-TAB}$ y $FD_{ZAN-CAJ}$. Así, se recomienda que tanto los examinadores expertos como noveles o con escaso nivel de entrenamiento con las principales pruebas de valoración del ROM_{FDT} utilicen la prueba $FD_{ZAN-INC}$ (figura 2) para: (a) identificar a personas con restringida o normal movilidad; y (b) monitorizar la eficacia de los programas de intervención aplicados en caso de ser necesarios.

Bibliografía

1. American Academy of Orthopaedic Association (1965). *Joint Motion: Method of Measuring and Recording*. Chicago: Park Ridge.
2. American Medical Association (2001). *Guides to the evaluation of permanent impairment*. 4th ed. Milwaukee, WI: Author.
3. Atkinson, G., y Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26(4), 217-238.
4. Atkinson, G., Nevill, A.M., y Edwards, B. (1999). What is an accep-

- table amount of measurement error? The application of meaningful "analytical goals" to the reliability of sports science measurements made on a ratio scale. *Journal of Sports Sciences*, 17, 18.
5. Beazell, J.R., Grindstaff, T.L., Sauer, L.D., Magrum, E.M., Ingersoll, C.D., y Hertel, J. (2012). Effects of a proximal or distal tibiofibular joint manipulation on ankle range of motion and functional outcomes in individuals with chronic ankle instability. *Journal of Orthopaedic & Sports Physical Therapy*, 42(2), 125-134.
 6. Bennell, K., Talbot, R., Wajswelner, H., Techovanich, W., Kelly, D., y Hall, A.J. (1998). Intra-rater and inter-rater reliability of a weight-bearing lunge measure of ankle dorsiflexion. *Australian Journal of Physiotherapy*, 44(3), 175-180.
 7. Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307-310.
 8. Bohannon, R.W., Tiberio, D., y Waters, G. (1991). Motion Measured from Forefoot and Hindfoot Landmarks During Passive Ankle Dorsiflexion Range of Motion 1. *Journal of Orthopaedic & Sports Physical Therapy*, 13(1), 20-22.
 9. Cejudo, A., Sainz de Baranda, P., Ayala, F., y Santonja, F. (2014). A simplified version of the weight-bearing ankle lunge test: Description and test-retest reliability. *Manual Therapy*, 19, 355-359
 10. Cejudo, A., Sainz de Baranda, P., Ayala, F., y Santonja, F. (2014). Test-retest reliability of seven common clinical tests for assessing lower extremity muscles flexibility in futsal and handball players. *Physical Therapy in Sport*. (in press).
 11. Denegar, C.R., Hertel, J., y Fonseca, J. (2002). The effect of lateral ankle sprain on dorsiflexion range of motion, posterior talar glide, and joint laxity. *Journal of Orthopaedic & Sports Physical Therapy*, 32(4), 166-173.
 12. DiGiovanni, C.W., Kuo, R., Tejwani, N., Price, R., Hansen Jr, S.T., Cziernecki, J., y Sangeorzan, B.J. (2002). Isolated gastrocnemius tightness. *Journal of Bone & Joint Surgery*, 84(6), 962-970.
 13. Grindstaff, T.L., Hertel, J., Beazell, J.R., Magrum, E.M., y Ingersoll, C.D. (2009). Effects of lumbopelvic joint manipulation on quadriceps activation and strength in healthy individuals. *Manual therapy*, 14(4), 415-420.
 14. Hoch, M. C., Staton, G. S., y McKeon, P. O. (2011). Dorsiflexion range of motion significantly influences dynamic balance. *Journal of Science and Medicine in Sport*, 14(1), 90-92.
 15. Hopkins, W.G. (2007). A spreadsheet for deriving a confidence interval, mechanistic inference and clinical inference from a p value. *Sports Science* 11, 16-20
 16. Hopkins, W.G. (2000a) Measures of validity. Sports science. <http://www.sportsci.org/index.html>. Accessed June 8, 2013.
 17. Hopkins, W.G. (2000b). Measures of reliability in sports medicine and science. *Sports Medicine*, 30(1), 1-15.
 18. Hopkins, W.G. (2004). Bias in Bland-Altman but not regression validity analyses. *Sports science*, 8(4).
 19. Hopkins, W.G., Marshall, S., Batterham, A., y Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine & Science in Sports & Exercise*, 41(1), 3-13.
 20. Kaufman, K.R., Brodine, S.K., Shaffer, R.A., Johnson, C.W., y Cullison, T.R. (1999). The effect of foot structure and range of motion on musculoskeletal overuse injuries. *American Journal of Sports Medicine*, 27(5), 585-593.
 21. Konor, M.M., Morton, S., Eckerson, J.M., y Grindstaff, T.L. (2012). Reliability of three measures of ankle dorsiflexion range of motion. *International Journal of Sports Physical Therapy*, 7(3), 279-287
 22. Krause, D. A., Cloud, B. A., Forster, L. A., Schrank, J. A., y Hollman, J. H. (2011). Measurement of ankle dorsiflexion: a comparison of active and passive techniques in multiple positions. *Journal of Sport Rehabilitation*, 20(3), 333-344.
 23. Landis, J.R., y Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159-174.
 24. Magee, D.J. (2002). *Orthopedic physical assessment*, (4th ed.), vol. 11. W.B. Saunders Company: Philadelphia, Pennsylvania.
 25. Martin, R.L., y McPoil, T.G. (2005). Reliability of ankle goniometric measurements: a literature review. *Journal of the American Podiatric Medical Association*, 95(6), 564-572.
 26. Mayer, T.G., Kondraske, G., Beals, S.B., y Gatchel, R.J. (1997). Spinal range of motion: accuracy and sources of error with inclinometric measurement. *Spine*, 22(17), 1976-1984.
 27. Munteanu, S.E., Strawhorn, A.B., Landorf, K.B., Bird, A.R., y Murley, G.S. (2009). A weightbearing technique for the measurement of ankle joint dorsiflexion with the knee extended is reliable. *Journal of Science and Medicine in Sport*, 12(1), 54-59.
 28. O'Shea, S., y Grafton, K. (2013). The intra and inter-rater reliability of a modified weight-bearing lunge measure of ankle dorsiflexion. *Manual Therapy*, 18(3), 264-268.
 29. Palmer, M.L. y Epler, M.E. (2002). *Fundamentos de las Técnicas de la Evaluación Musculoesquelética*. Barcelona: Paidotribo.
 30. Pope, R., Herbert, R., y Kirwan, J. (1998). Effects of ankle dorsiflexion range and pre-exercise calf muscle stretching on injury risk in army recruits. *Australian Journal of Physiotherapy*, 44(3), 165-172.
 31. Prentice, W.E. (2003). *The Thigh, Hip, Groin, and Pelvis*. In: *Arnheim's Principles of Athletic Training: A Competency-Based Approach* (11th ed). New York, NY: McGraw Hill.
 32. Scharf-billig, R., y Scutter, S. D. (2004). Measurement of foot dorsiflexion: a modified Lidcombe template. *Journal of the American Podiatric Medical Association*, 94(6), 573-577.
 33. Simondson, D., Brock, K., y Cotton, S. (2012). Reliability and smallest real difference of the ankle lunge test post ankle fracture. *Manual Therapy*, 17(1), 34-38.
 34. Stefanyshyn, D.J., y Engsborg, J.R. (1994). Right to left differences in the ankle joint complex range of motion. *Medicine and Science in Sports and Exercise*, 26(5), 551-555.
 35. Tiberio, D., Bohannon, R.W., y Zito, M.A. (1989). Effect of subtalar joint position on the measurement of maximum ankle dorsiflexion. *Clinical Biomechanics*, 4(3), 189-191.
 36. Venturini, C., Ituassú, N.T., Teixeira, L.M., y Deus, C.V.O. (2006). Intrarater and interrater reliability of two methods for measuring the active range of motion for ankle dorsiflexion in healthy subjects. *Brazilian Journal of Physical Therapy*, 10(4), 407-411.
 37. Vicenzino, B., Branjerdporn, M., Teys, P., y Jordan, K. (2006). Initial changes in posterior talar glide and dorsiflexion of the ankle after mobilization with movement in individuals with recurrent ankle sprain. *Journal of Orthopaedic & Sports Physical Therapy*, 36(7), 464-471.
 38. Weir, J.P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research*, 19(1), 231-240.

