



Fiabilidade da tela de movimento funcional: uma revisão sistemática

Reliability of the functional movement screen: a systematic review

Autores

João Vitor de Souza Moreira¹
Matheus Grandeni de Oliveira Pires¹
João Guilherme Vieira da Silva¹
Guilherme de Jesus do Nascimento¹
Camille dos Reis Molina¹
Jefferson Macedo Vianna¹
Jefferson da Silva Novaes¹
José Vilaça-Alves²
Aline Aparecida de Souza Ribeiro¹
André Calil e Silva¹

¹ Universidade Federal de Juiz de Fora (Brasil)

² Universidade de Trás-os-Montes e Alto Douro

Autor de correspondência:
João Vitor de Souza Moreira
joão200865@hotmail.com

Recibo: 22-05-25
Aceito: 20-04-26

Cómo citar na APA

de Souza Moreira, J. V., Grandeni de Oliveira Pires, M., Vieira da Silva, J. G., de Jesus do Nascimento, G., dos Reis Molina, C., Macedo Vianna, J., da Silva Novaes, J., Vilaça-Alves, J., Aparecida de Souza Ribeiro, A., & Calil e Silva, A. (2026). Fiabilidade da tela de movimento funcional: uma revisão sistemática. *Retos*, 78, 1250-1266. <https://doi.org/10.47197/retos.v80.116328>

Resumo

Introdução: Testes de triagem de movimento são utilizados como ferramentas para avaliar a capacidade de movimentação de indivíduos em determinadas situações. O Functional Movement Screen (FMS) é um teste de triagem composto por uma bateria de testes funcionais, empregados na avaliação de padrões considerados fundamentais.

Objetivos: Avaliar a confiabilidade intra-avaliador e inter-avaliador dos valores totais do Functional Movement Screen (FMS).

Metodologia: Conduziu-se uma revisão sistemática, seguindo as diretrizes do PRISMA, em bases de dados relacionadas à análise de movimento com o intuito de reunir fontes primárias sobre a confiabilidade do FMS. As bases de dados PubMed/MEDLINE, Embase, Web of Science e SPORTDiscus foram consultadas no período de 16/06/2024 a 26/10/2024. Foram incluídos estudos que reportaram a confiabilidade intra ou inter avaliador dos valores totais do FMS. **Resultados:** 13 artigos foram inclusos. Os valores de Coeficiente de Correlação Intraclasse (CCI) para confiabilidade inter-avaliador variaram entre 0,76 e 0,96, classificados como excelentes. Para a confiabilidade intra-avaliador, a maioria dos estudos reportou CCIs superiores a 0,75, também excelentes. A avaliação por vídeo e a presencial apresentaram resultados similares de confiabilidade. A experiência dos avaliadores não se mostrou um fator determinante. O uso de dois avaliadores demonstrou ser suficiente para excelente confiabilidade. A qualidade metodológica foi baixa, com apenas um artigo atingindo percentual superior a 60% na escala QAREL. **Conclusão:** Os resultados demonstram tendências que podem indicar o FMS como uma ferramenta de boa confiabilidade inter-avaliador e intra-avaliador, apesar das limitações metodológicas.

Palavras-chave

Confiabilidade; desempenho físico funcional; functional movement screen; reprodutibilidade dos testes.

Abstract

Introduction: Movement screening tests are used as tools to assess individuals' movement capacity in specific situations. The Functional Movement Screen (FMS) is a screening test composed of a battery of functional tests employed to evaluate movement patterns considered fundamental.

Objectives: To evaluate the intra-rater and inter-rater reliability of the total scores of the Functional Movement Screen (FMS).

Methodology: A systematic review was conducted following PRISMA guidelines in databases related to movement analysis, with the aim of gathering primary sources on the reliability of the FMS. The PubMed/MEDLINE, Embase, Web of Science, and SPORTDiscus databases were searched between June 16, 2024, and October 26, 2024. Studies reporting intra- or inter-rater reliability of the total FMS scores were included.

Results: Thirteen articles were included. Intraclass Correlation Coefficient (ICC) values for inter-rater reliability ranged from 0.76 to 0.96, classified as excellent. For intra-rater reliability, most studies reported ICCs above 0.75, also classified as excellent. Video-based and in-person assessments showed similar reliability results. Evaluator experience was not shown to be a determining factor. The use of two raters proved sufficient to achieve excellent reliability. Methodological quality was low, with only one article achieving a score above 60% on the QAREL scale.

Conclusion: The results demonstrate trends that may indicate the FMS as a tool with good inter-rater and intra-rater reliability, despite methodological limitations.

Keywords

Functional movement screen, physical functional performance, reliability, reproducibility of results.

Introdução

Os testes de triagem do movimento são ferramentas utilizadas para avaliar a capacidade do indivíduo executar padrões de movimentos globais ou específicos e, a partir deles, observar questões como equilíbrio, postura e movimentos articulares (McCunn et al., 2018). Estes testes podem ser utilizados com o intuito de diagnóstico clínico para a presença de dor ou dificuldade na realização do movimento avaliado (McCunn et al., 2018). Portanto, são considerados como uma ferramenta de auxílio para o profissional da área da saúde do movimento, seja ele fisioterapeuta ou profissional de Educação Física (Morgan et al., 2023). Os prognósticos, oriundos da utilização de tal ferramenta, podem ser usados em ambientes onde o movimento acontece de forma constante e sua eficiência é um determinante do sucesso do indivíduo, em atividades laborais ou esportivas (Moran et al., 2021). Uma aplicação específica dos testes de triagem dentro do cenário esportivo envolve a possibilidade de classificar os movimentos quanto a qualidade de execução e indicar procedimentos que possam contribuir para o aperfeiçoamento de técnicas esportivas e para a prevenção e tratamento de lesões (Morgan et al., 2023).

Para avaliar os padrões de movimentos no ambiente esportivo existem alguns testes de triagem com boa confiabilidade e que tem boa associação com o risco de lesão. Os mais utilizados são: Star Excursion Balance Test (SEBT), o Landing Error Scoring System, o Soccer Injury Movement Screen (SIMS), o Golf Movement Screen (GFS) e o Functional Movement Screen (FMS) (McCunn et al., 2016; McCunn et al., 2018; Gould et al., 2021). No que tange a predição de lesão e análise da qualidade de movimento, o FMS figura entre os mais utilizados por pesquisadores (McCunn et al., 2015; Silva et al., 2019; Asgari et al., 2021; Santana et al., 2021). Este instrumento foi foco de revisões sistemáticas que buscaram evidenciar sua validade, a relação com o desempenho atlético, a predição de lesão e a confiabilidade (Bonazza et al., 2017; Moran et al., 2017; Davies et al., 2022). Investigações que buscam verificar a confiabilidade de um instrumento pretendem demonstrar a reprodutibilidade da ferramenta, já que a confiabilidade é fundamental para demonstrar a precisão da medida e da variação dessa medida, tanto no âmbito das pesquisas quanto no campo prático (Hopkins, 2000; McCann et al., 2015).

Dada essa importância, ainda que uma gama de investigações teve como foco a confiabilidade do FMS, os valores encontrados divergem entre os estudos. Uma excelente confiabilidade foi encontrada nos estudos de Schneiders et al. (2011), Parenteau-G et al. (2013) e Morgan et al. (2023). Níveis de confiabilidade com valores bons foram encontrados por Gribble et al. (2013) e Shultz et al. (2013) e mais recentemente Ab Malik et al. (2025) encontraram valores ruins de confiabilidade. Diante deste quadro de divergência na literatura científica pode-se demonstrar a importância de revisões que sumarizem os resultados para um melhor conhecimento sobre a confiabilidade do instrumento. Dentre as revisões sistemáticas que trataram da confiabilidade do FMS (Cuchna et al., 2016; Moran et al., 2016; Bonazza et al., 2017), nenhuma delas foi capaz de elencar todos os procedimentos de avaliação do teste que influenciam sua confiabilidade.

Cuchna et al. (2016) não comparam os valores de confiabilidade obtidos a partir da avaliação realizada em tempo real com os valores obtidos a partir de uma avaliação por vídeo. Portanto, deixam a dúvida sobre qual forma de avaliar seria mais confiável. Apesar de Moran et al. (2016) realizarem a comparação citada anteriormente, não discutem a influência do nível de experiência dos avaliadores. Ainda, Bonazza et al. (2017) que tiveram como foco da revisão investigar a confiabilidade, validade e predição de lesão do FMS, não trataram das questões relacionadas à forma de avaliação (tempo real x avaliação por vídeo) e o nível de experiência dos avaliadores. Ademais, é possível justificar a necessidade de uma nova revisão a partir de estudos que não foram incluídos em nenhuma dessas revisões (Smith et al., 2017; Harper e Glass, 2020; Morgan et al., 2023) de forma que novas conclusões podem ser tomadas a partir desses artigos publicados no decorrer de oito anos desde a última revisão conhecida sobre o tema.

Diante deste contexto, emerge a necessidade de se realizar uma revisão sobre a confiabilidade do FMS que responda às seguintes questões: (a) quais os valores de confiabilidade inter e intra-avaliador do FMS; (b) qual a diferença dos valores de confiabilidade obtidos a partir da avaliação por tempo real e por vídeo; (c) qual o nível ideal de experiência dos avaliadores que pode interferir na confiabilidade do FMS; (d) como a influência do qual o número de avaliadores e o número de avaliados que pode influenciar na pode afetar a confiabilidade do FMS. Para responder essas questões, o objetivo principal da revisão sistemática com meta-análise é avaliar a confiabilidade intra-avaliador e inter-avaliador dos valores totais do Functional Movement Screen (FMS). Como objetivos secundários, a revisão visa responder quais as



diferenças entre os valores de confiabilidade obtidos a partir da visualização por tempo real e por vídeo, como a experiência dos avaliadores afeta a confiabilidade do FMS e qual a influência do número de avaliadores e avaliados na confiabilidade do FMS.

Método

Protocolo e Registro

Essa revisão sistemática seguiu os moldes do Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Liberati et al., 2009; Shamseer et al., 2015; Page et al., 2021; Molina et al., 2025) e foi registrada na Open Science Framework (OSF) com DOI de número: 10.17605/OSF.IO/7NHCP. Os métodos de revisão foram estabelecidos antes do início da pesquisa.

Critérios de elegibilidade

Os artigos foram incluídos a partir da leitura dos títulos, resumos e textos completos e atenderam aos seguintes critérios:

Critérios de inclusão

1. Artigos que foram publicados em língua inglesa.
2. Artigos que trataram da confiabilidade do FMS em algum de seus desfechos.
3. Artigos que foram publicados em periódicos revisados por pares.
4. Artigos originais com texto completo.
5. Artigos que utilizaram o Índice de Correlação Intraclasse (CCI) para a avaliação da confiabilidade do somatório dos testes do FMS.

Critérios de exclusão

1. Artigos que aplicaram o FMS em indivíduos com lesões musculoesqueléticas ou enfermidade que comprometessem a realização de algum dos testes.
2. Artigos que não foram claros sobre o índice estatístico usado no cálculo da confiabilidade.
3. Artigos que mencionaram a confiabilidade realizada em estudo anterior.
4. Artigos de revisão, teses, dissertações e resumos de congresso.
5. Artigos que não avaliaram a confiabilidade do valor total do FMS.

Estratégia de busca

As buscas aconteceram no período de 26/06/2024 a 16/10/2024 nas bases de dados PubMed; Embase; Web of Science e SPORTDiscus com a busca conduzida pelos pesquisadores JVSM e MGSP. Essas bases de dados foram escolhidas por abrigarem artigos relacionados a testes clínicos e a análise de movimento. Diferentes estratégias de busca foram aplicadas em cada uma das bases.

Para determinar as estratégias de busca, uma revisão preliminar da literatura foi realizada a fim de identificar termos comuns relacionados a esta pesquisa. Essas buscas resultaram na inclusão das seguintes palavras-chave: "functional movement screen", "fms", "reliability", "reproducibility", "inter-rater reliability", "intra-rater reliability", "expert", "experienced", "novice", "beginner", "individual tests", "test components", "seven tests", "video assessment", "real-time assessment" e "live assessment". As buscas dos termos foram combinadas usando dois operadores Booleanos (AND/OR).

Processo de seleção dos estudos

Os estudos foram recolhidos e catalogados a partir do software Rayyan. Um aplicativo que existe para facilitar o processo de triagem dos artigos, além de permitir o trabalho cooperativo de forma simultânea, disponível em: <http://rayyan.qcri.org> (acessado em 26/10/2024) (Ouzzani et al., 2016). Os artigos duplicados foram removidos de forma automática e manual (JVSM e MGSP).



Processo de coleta dos dados

Os dados relevantes foram definidos de maneira prévia pelos dois pesquisadores (JVSM e MGSP). As discordâncias apresentadas em algum processo da triagem que foi feita de maneira independente e foram resolvidas pelo terceiro revisor(ACS), a partir da análise dos argumentos e critérios utilizados pelos revisores primários. Os dados foram posteriormente recolhidos e analisados em planilhas do Excel.

Dados coletados

Os seguintes dados foram extraídos das fontes primárias: autores e ano de publicação, desenho da confiabilidade (intra ou inter-avaliador), número dos avaliadores, experiência dos avaliadores, técnica de visualização, valores de confiabilidade total intra e inter-avaliador e amostra(participantes). Essas características foram expostas de maneira descritiva na tabela 2 que sintetizaram todos os dados.

Estudo do risco de viés

A escala QAREL (Quality Appraisal of Reliability Studies) foi usada para determinar o risco de viés dos artigos, disponível para visualização na figura 1 (Lucas et al., 2010). Essa escala foi criada para avaliar a qualidade metodológica de estudos que mostram diagnósticos de confiabilidade, incluindo testes de triagem de movimento como o FMS. Ela investiga questões como os métodos estatísticos utilizados, a validade do teste, a quantidade de informações disponíveis aos avaliadores, as formas de aplicação e a interpretação dos testes. O instrumento é formado por onze questões que são respondidas com sim, não, inconclusivo ou não aplicável(N/A), onde cada uma das perguntas é avaliada com o mesmo peso. Cada resposta que foi dada como “sim” foi somada e a porcentagem de cada uma das questões foi calculada com base na resposta dos dois pesquisadores (JVSM e MGSP) que responderam a escala de maneira independente. Discordâncias na resposta foram resolvidas por um especialista (ACS).

Apesar da escala QAREL não mencionar um corte percentual para definir estudos como de alta qualidade metodológica (Lucas et al. 2010), diversos artigos de revisão de literatura a respeito de confiabilidade que utilizaram esse instrumento estabeleceram um percentual igual ou maior que 60% de respostas “sim”, como um critério para alta qualidade metodológica (May; Littlewood; Bishop, 2006; May et al., 2010; Lucas et al., 2010; Cuchna et al., 2016; Moran et al . 2016). Assim, adotou-se este critério para esta pesquisa.

Análise de dados e métodos de síntese

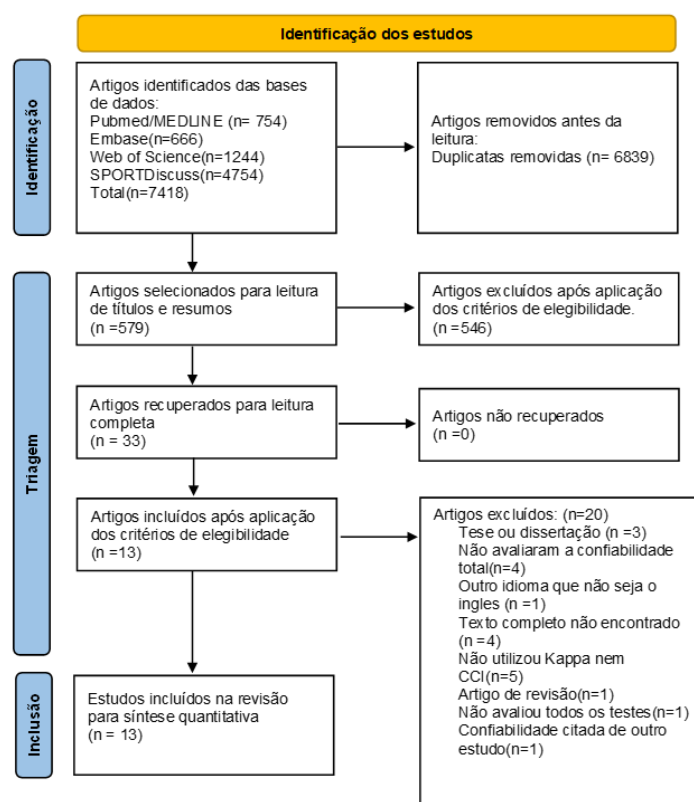
Foram realizadas diferentes metá- análises para cada um dos índices de confiabilidade, intra- avaliador e inter-avaliador. Foram utilizadas análises com diferentes subgrupos, desde que fosse possível agrupar pelo menos 3 estudos (Jackson;Turner, 2017). Os subgrupos foram: 1) visualização por vídeo e por tempo real; 2) avaliadores experientes e não experientes, tanto do índice inter-avaliador quanto do índice intra-avaliador. Além disso, foram realizadas meta-análises utilizando o número de avaliadores e o número de avaliados como moderadores. Alguns estudos não foram incluídos em ambas as análises devido ao fato de não investigarem as duas formas de confiabilidade simultaneamente. Para cada metá-análise, o número da amostra e os valores de CCI foram extraídos de cada estudo. Como a variância de cada CCI não foi diretamente reportada, o intervalo de confiança foi utilizado para determinar a variância com o método Fisher (Bonazza et al., 2017). Uma análise de heterogeneidade foi utilizada com o índice estatístico I^2 (Migliavaca et al., 2022). Todas essas análises foram realizadas no software Jamovi (versão 2.3.8) utilizando a extensão Major. Foi realizada uma análise de sensibilidade a partir do cálculo de CCI com a retirada de cada um dos estudos, visando observar a presença de outliers nos artigos.

Resultados

Na fase de identificação, foram encontrados 7418 artigos durante as buscas nas bases de dados usando as palavras-chaves listadas. Foram removidas as duplicatas que totalizaram 6839 artigos. Na fase de triagem, após a leitura de títulos e resumos, 579 artigos foram identificados. Com a aplicação dos critérios de elegibilidade, foram removidos 546 artigos. Assim, 33 artigos foram selecionados para a leitura completa. Na fase de inclusão, foram verificados os critérios de elegibilidade e foram inclusos 13 artigos para síntese quantitativa (ver figura 1 com o Fluxograma PRISMA 2020).



Figura 1. Fluxograma PRISMA 2020 mostrando o número de artigos reportados e o número de artigos elegíveis após o processo de busca.



*Fonte: Elaborado pelos autores (2025)

Classificação da Confiabilidade

A confiabilidade foi classificada, qualitativamente, baseada nos critérios de Cicchetti (1994). Segundo esse autor, valores abaixo ou iguais a 0.4 são considerados ruins, de 0.4 e 0.59 são considerados razoáveis, entre 0.6 e 0.74 são valores bons e de 0.75 a 1 é considerado um valor excelente.

Tabela 1. Classificação da confiabilidade.

Valores de CCI	Cicchetti (1994)
≤ 0.4	Ruim
0.4-0.59	Razoável
0.6-0.75	Boa
0.75-1.00	Excelente

Descrição dos Resultados

Em relação à avaliação metodológica dos estudos, os resultados da verificação do risco de viés podem ser observados na tabela 2.

Tabela 2. Risco de viés dos estudos

Questões	Schneiders et al. (2011)	Teyhen et al. (2012)	Gribble et al. (2013)	Maeda et al. (2013)	Smith et al. (2013)	Shultz et al. (2013)	Parenteau-G. et al. (2014)	Gulgin (2014)	Song et al. (2014)	Saki et al. (2016)	Smith et al. (2017)	Harper e Glass (2020)	Morgan et al. (2023)
1. Amostra representativa	Sim	Sim	Inconclusivo	Inconclusivo	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Inconclusivo	Inconclusivo
2. Avaliadores representativos	Inconclusivo	Sim	Sim	Sim	Sim	Sim	Inconclusivo	Sim	Inconclusivo	Sim	Inconclusivo	Sim	Sim
3. Blinding (avaliadores)	Sim	Sim	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Sim	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Não	Sim
4. Blinding (resultados)	N/A	Sim	Inconclusivo	Inconclusivo	Inconclusivo	Sim	Inconclusivo	N/A	N/A	Inconclusivo	Inconclusivo	N/A	Inconclusivo
5. Blinding (referência/doença)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
6. Viés clínico	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	N/A	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo
7. Informações adicionais	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo	Inconclusivo
8. Ordem de exame variada	Inconclusivo	Sim	Sim	Inconclusivo	Não	Não	Sim	Não	Inconclusivo	Inconclusivo	Sim	Inconclusivo	Inconclusivo
9. Intervalo de tempo apropriado	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
10. Teste apropriado	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
11. Estatística apropriada	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Inconclusivo	Sim	Sim	Sim	Não
12. Validade interna (%)	28,6%	57,1%	28,6%	14,3%	14,3%	14,3%	42,8%	14,3%	14,3%	14,3%	28,6%	14,3%	28,6%
13. Validade externa (%)	100%	100%	66,7%	66,7%	100%	100%	66,7%	100%	66,7%	100%	66,7%	66,7%	66,7%
14. Método estatístico (%)	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	0%
15. % geral (Sim)	45,6%	72,8%	45,6%	36,4%	45,6%	54,5%	54,6%	45,6%	27,3%	45,6%	45,6%	36,4%	36,4%

** N/A= Não aplicável

Na tabela 2, em relação a quantidades de respostas como sim, referente ao item 15 da figura acima, os resultados mais notáveis foram que: (1) nenhum artigo apresentou qualidade metodológica para a validade interna. Verificou-se que o maior percentual relativo de validade interna foi de 57,1 % (Teyhen et al., 2012); (2) todos os artigos apresentaram alta qualidade metodológica para a validade externa, com percentual de respostas “sim” maiores que 60%; (3) 11 dos 13 artigos apresentaram qualidade para os métodos estatísticos utilizados como avaliação de confiabilidade. Notou-se que apenas um estudo (Morgan et al., 2023) não apresentou qualidade; (4) em relação ao percentual total de respostas “sim”, para as onze questões do instrumento da qualidade metodológica, somente um artigo (Teyhen et al., 2012) apresentou valor igual ou superior a 60%, com 72,8%.

Avaliando os valores de confiabilidade intra e inter-avaliador do FMS, a tabela 3 apresenta 7 dos 13 artigos com desenho de confiabilidade intra e inter avaliador, de forma simultânea. A confiabilidade inter-avaliador estava presente em 5 artigos. Apenas 1 estudo investigou somente a confiabilidade intra-avaliador.

Sobre a avaliação em tempo real e por vídeo, os estudos que investigaram a confiabilidade intra e inter-avaliador apresentaram um quantitativo de 4 para a visualização em tempo real, de 1 para a visualização por vídeo e 2 para ambas as formas de visualização. Os estudos que investigaram apenas a confiabilidade inter-avaliador apresentaram um quantitativo de 2 para a visualização em tempo real e 3 para a visualização por meio de vídeo. O estudo que investigou somente a confiabilidade intra-avaliador utilizou o vídeo como forma de visualização.

A experiência dos avaliadores nos estudos que investigaram os dois índices (intra e inter-avaliador) varia no quantitativo de 1 para 31 avaliador considerados como novatos e de 1 a 7 para os avaliadores

considerados como experientes. Os estudos que investigam somente a confiabilidade inter-avaliador apresentam de 2 a 4 avaliadores novatos e de 1 a 2 avaliadores experientes. O estudo que investiga somente a confiabilidade intra-avaliador apresenta 31 avaliadores inexperientes e 7 avaliadores experientes.

Para o número de avaliadores notamos que o quantitativo varia de 3 a 8 avaliadores para os estudos que investigam a confiabilidade intra e inter-avaliador. Nos estudos que investigaram a confiabilidade inter-avaliador verificou-se de 3 a 5 avaliadores. Para o estudo que investiga somente a confiabilidade intra-avaliador o número de avaliadores foi de 38.

Sobre os números de participantes (avaliados) os estudos que investigaram as confiabilidades intra e inter-avaliador apresentaram amostras que continham de 12 a 94 participantes. Os estudos que investigaram apenas a confiabilidade inter-avaliador apresentaram um quantitativo que variou de 200 a 16 participantes. O artigo que investigou apenas a confiabilidade intra-avaliador teve uma amostra de 3 participantes.

Tabela 3. Características dos estudos incluídos.

Autor	Confiabilidade total inter-avaliador	Confiabilidade total Intra-avaliador	Nível de Confiabilidade (Segundo Ciccheci,1994)	Técnica de visualização	Experiência dos avaliadores	Número de Avaliadores	Número de avaliados
Schneiders et al. (2011)	CCI =0.971	Não reportado	Excelente	Tempo real	2 inexperientes	2	200
Teyhen et al. (2012)	CCI =0.76	CCI = 0.74	Inter=Excelente Intra=Boa	Tempo real	8 inexperientes	8	64
Gribble et al. (2013)	Não reportado	CCI1=0.946 CCI2=0.758 CCI3=0.372	Intra1=Excelente Intra2=Excelente Intra3= Ruim	Vídeo	7 experientes e 31 inexperientes	38	3
Maeda et al. (2013)	Não reportado	CCI =0.95	Excelente	Tempo real	1 avaliador inexperiente e 1 experiente	1 intra e 2 inter	12
Shultz et al. (2013)	Não reportado	CCI=0.6	Boa	Tempo real	5 treinadores experientes e 1 novato	1 intra e 5 inter	19
Smith et al. (2013)	CCI1=0.89 CCI2=0.87	CCI1=0.900 CCI2=0.810 CCI3=0.910 CCI4= 0.888	Excelente	Tempo real	4 inexperientes	4	19
Parenteau-g et al. (2014) Gulgin;	CCI=0.96	CCI=0.96	Excelente	Vídeo	4 experientes	2 intra e 2 inter	28
Hoogenbomm (2014)	CCI=0.882	Não reportado	Excelente	Vídeo	3 novatos e 1 experiente	4	20
Saki(2016)	CCI=0.96	CCI=0.87	Excelente	Vídeo	1 experiente e 2 novatos	3 inter e 1 intra	15
Smith et al. (2017)	CCI=0.88	CCI=0.83	Excelente	Tempo real	4 experientes	4	94
Harper & Glass (2020)	CCI1=0.79 CCI2=0.84	Não reportado	Excelente	Tempo real	1 experiente e 4 novatos	5	16
Morgan et al. (2023)	CCI=0.95	Não reportado	Excelente	Vídeo	4 inexperientes	4	45

*Intra= intra-avaliador Inter=inter-avaliador CCI=Coeficiente de correlação intraclasse.

Confiabilidade Inter-Avaliador

A meta- análise realizada para confiabilidade inter-avaliador demonstrou um CCI agrupado de 0.90 [0.86, 0.94] conforme demonstrado na figura 1 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos foi elevada, com um I^2 de 86.53% (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 2 (Apêndice 1).

Confiabilidade Intra- Avaliador

A meta- análise realizada para confiabilidade intra-avaliador demonstrou um CCI agrupado de 0.88 [0.84, 0.93] conforme demonstrado na figura 3 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos foi elevada,



com um I^2 de 75.83%, (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 4 (Apêndice 1).

Forma de visualização da confiabilidade intra-avaliador

A meta-análise realizada para a confiabilidade intra-avaliador nos estudos que realizaram a visualização em tempo real demonstrou um CCI agrupado de 0.86[0.81, 0.92] conforme demonstrado na figura 5 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram o FMS em tempo real foi elevada, com um I^2 de 61.34% (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 6 (Apêndice 1).

A meta-análise que verificou a confiabilidade intra-avaliador dos estudos que utilizaram o vídeo como forma de visualização, demonstrou um CCI agrupado de 0.96 [0.93, 0.98] conforme demonstrado na figura 7 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram o FMS por vídeo foi baixa, com um I^2 de 0.46% (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 8 (Apêndice 1).

Forma de visualização da confiabilidade inter-avaliador

A meta-análise realizada para a confiabilidade inter-avaliador nos estudos que realizaram a visualização em tempo real demonstrou um CCI agrupado de 0.87[0.82, 0.93] conforme demonstrado na figura 9 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram o FMS na confiabilidade inter-avaliador em tempo real foi elevada, com um I^2 de 80.4%, (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 10 (Apêndice 1).

A meta-análise que verificou a confiabilidade inter-avaliador dos estudos que utilizaram o vídeo como forma de visualização, demonstrou um CCI agrupado de 0.95 [0.93, 0.97] conforme demonstrado na figura (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram a confiabilidade inter-avaliador do FMS por vídeo foi baixa, com um I^2 de 0.52%, (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura (Apêndice 1).

Nível de experiência na confiabilidade intra-avaliador

A meta-análise que verificou a confiabilidade intra-avaliador dos estudos que utilizaram avaliadores experientes no FMS demonstraram um CCI agrupado de 0.93 [0.88, 0.98] conforme demonstrado na figura 13 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram a confiabilidade intra-avaliador do FMS com avaliadores experientes foi alta, com um I^2 de 77.95%, (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 14 (Apêndice 1).

A meta-análise que verificou a confiabilidade intra-avaliador dos estudos que utilizaram avaliadores inexperientes no FMS demonstraram um CCI agrupado de 0.92 [0.87, 0.96] conforme demonstrado na figura 15 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram a confiabilidade intra-avaliador do FMS com avaliadores inexperientes foi alta, com um I^2 de 77.95%, (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 16 (Apêndice 1).

Nível de experiência na confiabilidade inter-avaliador

A meta-análise que verificou a confiabilidade inter-avaliador dos estudos que utilizaram avaliadores inexperientes no FMS demonstraram um CCI agrupado de 0.90 [0.83, 0.97] conforme demonstrado na figura 19 (Apêndice 1). Esse valor indica uma confiabilidade excelente segundo Cicchetti (1994). A análise de heterogeneidade encontrada entre os estudos que verificaram a confiabilidade inter-avaliador do FMS com avaliadores inexperientes foi alta, com um I^2 de 77.95%, (Migliavaca et al., 2022). As estatísticas de heterogeneidade podem ser observadas na figura 20 (Apêndice 1). Gulgin;Hoogenboon (2014), Saki (2016) e Harper e Glass(2023) não entraram nessa meta-análise por alocarem avaliadores experientes e inexperientes para o mesmo cálculo de CCI.

Número de Avaliadores

A meta-análise que verificou a influencia do número de avaliadores na confiabilidade intra-avaliador indica que um menor número de avaliadores é um indicativo de melhor confiabilidade. Como pode ser observada na figura 21 (Apêndice 1).

A meta-análise que verificou a influencia do número de avaliadores na confiabilidade inter-avaliador indica que um menor número de avaliadores é um indicativo de melhor confiabilidade. Como pode ser observada na figura 22 (Apêndice 1).

Número de Avaliados

A meta-análise que verificou a influencia do número de avaliados na confiabilidade intra-avaliador indica que um maior número de avaliados é um indicativo de melhor confiabilidade. Como pode ser observada na figura 23 (Apêndice 1).

A meta-análise que verificou a influencia do número de avaliados na confiabilidade inter-avaliador indica que um maior número de avaliados é um indicativo de melhor confiabilidade. Como pode ser observada na figura 24 (Apêndice 1).

A meta-análise que verificou a influencia do número de avaliadores na confiabilidade inter-avaliador indica que um maior número de avaliados é um indicativo de melhor confiabilidade. Como pode ser observada na figura 24 do apêndice.

Análise de sensibilidade

A análise de sensibilidade inter-avaliador pode ser observada na figura 25. A análise dos dados ficou mais sensível com a retirada do artigo de Morgan et al. (2023), apresentando um I² de 0.

Figura 2. Análise de sensibilidade inter-avaliador.

Study	ICC 95%	I ² (%)
All	0,87 [0,78; 0,97]	71,52
Without Teyhen et al. (2012)	0,89 [0,80; 0,98]	56,62
Without Schneiders et al. (2011)	0,87 [0,77; 0,97]	60,44
Without Smith et al. (2013) ¹	0,88 [0,79; 0,97]	76,77
Without Smith et al. (2013) ²	0,88 [0,79; 0,97]	75,8
Without Parenteau-g et al. (2014)	0,87 [0,77; 0,97]	61,68
Without Gulgin; Hoogen-bomm (2014)	0,88 [0,79; 0,97]	76,5
Without Saki (2016)	0,88 [0,80; 0,97]	72,05
Without Smith et al. (2017)	0,88 [0,79; 0,97]	76,06
Without Harper & Glass (2020) ¹	0,89 [0,80; 0,97]	72,73
Without Harper & Glass (2020) ²	0,88 [0,80; 0,97]	74,19
Without Morgan et al. (2023)	0,87 [0,77; 0,97]	0

A análise de sensibilidade intra-avaliador pode ser observada na figura 26. A análise dos dados ficou mais sensível com a retirada do artigo de Saki et al. (2016).

Figura 2. Análise de sensibilidade intra-avaliador.

Study	ICC 95%	I ² (%)
All	0,82 [0,68; 0,95]	74,18
Without Teyhen (2012)	0,82 [0,68; 0,96]	63,62
Without Gribble(2013) ¹	0,85 [0,74; 0,96]	72,89
Without Gribble (2013) ²	0,82 [0,69; 0,95]	75,3
Without Gribble (2013) ³	0,81 [0,67; 0,95]	75,43
Without Maeda et al. (2013)	0,81 [0,66; 0,95]	75,23
Without Shultz et al. (2013)	0,83 [0,71; 0,96]	72,35
Without Smith et al. (2013) ¹	0,81 [0,67; 0,95]	78,13
Without Smith et al. (2013) ²	0,82 [0,68; 0,95]	74,65
Without Smith et al. (2013) ³	0,81 [0,67; 0,95]	78,06
Without Smith et al. (2013) ⁴	0,81 [0,67; 0,95]	77,93
Without Paranteau-g et al. (2014)	0,80 [0,66; 0,95]	71,23
Without Saki et al. (2016)	0,80 [0,66; 0,95]	60,78
Without Smith et al. (2017)	0,82 [0,67; 0,96]	75,5

Discusión

Confiabilidade inter-avaliador

Os valores apresentados de CCI ficaram entre 0.76 (Teyhen et al., 2012) e 0.96 (Parenteau-g et al., 2014; Saki, 2016). O CCI agrupado encontrado foi de 0.90 [0.86, 0.94]. Portanto, a confiabilidade apresentada se enquadra dentro do considerado excelente para os parâmetros de Cicchetti (1994). A análise de sensibilidade não demonstrou grande discrepância do valor de confiabilidade inter-avaliador total, com os resultados de CCI indo de 0.87[0.77,0.97] a 0.89[0.80,0.98]. Dessa forma, o resultado da confiabilidade total não ficou altamente sensível a nenhum estudo específico. Os resultados dentro desse intervalo indicam uma boa reprodutibilidade entre as diferentes avaliações. Este fato demonstra uma consistência na aplicação do teste, uma vez que padrões semelhantes são julgados de forma similar. A confiabilidade inter-avaliador do FMS se mostrou consistente em todos os estudos, mesmo com diferentes grupos amostrais, diferentes níveis de experiência e números diferentes de avaliadores. A pequena faixa de variação dos valores (0.76 a 0.96) parece evidenciar que os valores totais da confiabilidade inter-avaliador do FMS sofrem pouca alteração, mesmo quando aplicados por diferentes grupos de avaliadores em diferentes participantes (Bonazza et al., 2017). Tais resultados indicam que futuros estudos podem ser capazes de aplicar o FMS com diferentes tipos de avaliadores e diferentes tipos de amostra.

Confiabilidade intra-avaliador

Foram encontrados resultados divergentes sobre a confiabilidade intra-avaliador na aplicação do FMS. No entanto, a maioria dos artigos demonstrou excelente confiabilidade em pelo menos um de seus grupos. O CCI agrupado encontrado foi de 0.88 [0.84, 0.93] o que é excelente para os parâmetros de Cicchetti(1994). A análise de sensibilidade não demonstrou grande discrepância do valor de confiabilidade intra-avaliador total, com os resultados de CCI indo de 0.81[0.66,0.95] a 0.85[0.74,0.96]. Dessa forma, o resultado da confiabilidade intra-avaliador total não ficou altamente sensível a nenhum estudo específico. Dos sete estudos que investigaram o índice intra-avaliador em algum de seus desfechos, cinco (Maeda et al., 2013; Smith et al., 2013; Parenteau-g et al., 2014; Saki, 2016; Smith et al., 2017) apresentaram CCIs superiores a 0.75 em todos os seus grupos. Em contrapartida, Teyhen et al. (2012) encontraram CCI igual a 0.74. Valor que se aproxima do excelente e se apresenta no intervalo considerado como bom para Cicchetti (1994), dentro do corte de 0,6 até 0,74. Shultz et al., (2013) também apresentaram um valor considerado bom, com CCI de 0.6 para a confiabilidade intra-avaliador, possivelmente devido ao alto intervalo de confiança apresentado, o que dificulta a presença de um valor de CCI mais elevado. Dessa maneira, o estudo de Shultz et al.(2013) é um demonstrativo que pode existir uma alta variabilidade de resultados mesmo dentro de um cálculo que demonstre boa confiabilidade. Essa situação indica que alguns avaliadores podem ter dificuldade de obter resultados consistentes com suas próprias avaliações.

De uma maneira geral, os resultados foram favoráveis, entretanto Gribble et al. (2013) foi o único estudo desfavorável que avaliou apenas a confiabilidade intra-avaliador. Nele, houve uma demonstração de um

baixo nível de confiabilidade. Em seu estudo, os avaliadores foram divididos, de acordo com sua experiência com o FMS em três grupos diferentes. Dois grupos de avaliadores mais experientes apresentaram excelentes índices de confiabilidade intra-avaliador (CCI=0.946 e CCI=0.758), enquanto o terceiro grupo, formado por estudantes inexperientes na avaliação, apresentou um índice de confiabilidade ruim (CCI=0.372). Dois fatores devem ser considerados para esse resultado, o número de avaliadores que formaram esse grupo e o nível de experiência dos mesmos. Sobre o primeiro fator, o referido grupo apresentou o maior número de avaliadores, por grupo, dentre todos os estudos (dezesesseis indivíduos). Fato que potencializa a possibilidade de uma maior variabilidade de resultados. Sobre o diferente nível de experiência do terceiro grupo, composto por estudantes, foi observado que a confiabilidade foi reduzida significativamente conforme avaliadores sem prática com o FMS que foram alocados em um mesmo grupo. Isso é uma evidência que contradiz a crença demonstrada pelos criadores do FMS que diz que a experiência pouco influí na confiabilidade do teste.

Além disso, o índice intra-avaliador depende de uma reexibição do avaliado, sendo esta confiabilidade intra-avaliador afetada, possivelmente, pelo tempo entre uma visualização e outra. Tal aspecto carece de padronização entre os estudos, dificultando a reprodutibilidade metodológica dos estudos. Na metodologia de Teyhen et al., (2012), o tempo entre sessões foi de um dia, sendo a segunda sessão no dia subsequente à primeira. Quatro estudos utilizaram intervalo de uma semana entre a primeira e a segunda avaliação (Maeda et al., 2013; Gribble et al., 2013; Smith et al., 2013; Saki, 2016). Já Parenteau-g et al., (2014) espaçaram as sessões de visualização em seis semanas. Por último, Shultz et al. (2013) e Smith (2017) não mencionaram o intervalo utilizado entre os testes. Nesse sentido, infere-se que maiores intervalos entre a primeira e a segunda avaliação proporcionam um espaço maior para o esquecimento das pontuações atribuídas anteriormente, o que diminui o risco de viés. Os estudos futuros devem padronizar melhor o intervalo de teste e reteste, com preferência para os intervalos maiores.

Forma de visualização

As metodologias dos estudos apresentaram duas formas de avaliação, por gravações de vídeo e em tempo real. Dos treze artigos selecionados, sete utilizaram o vídeo em pelo menos um de seus grupos, com CCIs que variaram de 0.372 (Gribble et al., 2013) até os 0.96 (Parenteau-g et al., 2014). Notou-se que não há uma padronização da metodologia usada na filmagem dos testes, apesar de todos possibilitarem visão nos planos sagital e frontal e utilizarem pelo menos duas câmeras. A frequência e possibilidade de mais de uma visualização variou entre os artigos.

Em relação a confiabilidade intra-avaliador, os estudos que utilizaram o vídeo para análise da confiabilidade apresentaram um CCI agrupado de 0.96 [0.93, 0.98] demonstrando excelente confiabilidade. Sobre o índice intra-avaliador um aspecto importante é a diferença de tempo na avaliação teste-reteste. Uma vez que ela pode influenciar na medida em que o mesmo examinador lembra da nota anterior dada ao indivíduo, especialmente se o período de reexibição for curto, já que se trata de uma nova visualização dos mesmos vídeos. Nesse sentido, verificou-se um estudo que repetiu a exibição depois de um período de seis semanas (Parenteau-G et al., 2014), outro que repetiu a exibição depois de sessenta minutos (Smith et al., 2017) e estudos que avaliaram novamente depois de uma semana (Gribble et al., 2013; Shultz et al., 2016; Saki, 2016;). Dente esses estudos que utilizaram uma semana de intervalo, observa-se que dois estudos (Gribble et al., 2013; Shultz et al., 2016) apresentaram confiabilidade ruim e boa, com CCIs de 0.372 e 0.6 respectivamente. Estes utilizaram o mesmo intervalo de tempo de reexibição, o que pode ser um indicativo de qual o intervalo não se deve utilizar. Já em relação a confiabilidade inter-avaliador a partir da avaliação por vídeo, esta não parece ser uma variável que irá influenciar negativamente, uma vez que os valores encontrados foram excelentes em todos os estudos que aplicaram esse método com um CCI agrupado de 0.95 [0.93, 0.97] e sem a presença de nenhum outlier como ocorreu para a confiabilidade intra-avaliador, como por exemplo o índice ruim de Gribble et al. (2013) com valor de 0.372.

Para os estudos que utilizaram a visualização por tempo real relativo ao índice intra-avaliador o CCI agrupado encontrado foi de 0.86 [0.81, 0.92], onde sete a utilizaram em pelo menos um de seus grupos. Todos os estudos que aplicaram a visualização por tempo real apresentaram resultados de confiabilidade excelentes. O CCI agrupado da confiabilidade inter-avaliador para o tempo real foi de 0.87 [0.82, 0.93] com, um estudo apresentando um valor bom considerado para Ciccheci (1994). Teyhen et al., (2012) encontraram CCI de 0.74. Ao observarmos a diferença entre os valores de CCI para a visualização

por vídeo e tempo real para a confiabilidade inter-avaliador, podemos notar que o vídeo apresenta resultados melhores. Fato este que pode ser decorrente da maior instrução previa passada aos avaliados antes da filmagem dos vídeos, o que não ocorre em tempo real, uma vez que não há necessidade de realização do teste novamente.

Sobre o índice intra-avaliador, podemos dizer que de maneira geral a confiabilidade entre as avaliações presenciais e em vídeo apresentaram valores similares. Vale destacar que questões como os ângulos de filmagem e o tempo de intervalo entre as exposições são relevantes para a avaliação a partir de gravações, uma vez que os avaliadores somente terão estes ângulos para visualização dos testes. Uma possível vantagem desse método está relacionada a capacidade de refazer o exame caso ocorra algum problema, o que também facilita o cálculo do CCI intra-avaliadores. A escolha do vídeo pode ser uma facilitadora devido a capacidade de reprodução por quantas vezes for necessário, algo que não é possível na avaliação presencial. Ainda sobre vantagens e desvantagens, a preferência pelo uso do vídeo também pode estar relacionada a questões como a logística e a mobilidade de avaliados e avaliadores, que para a avaliação em tempo real necessitam do deslocamento para o mesmo espaço no mesmo horário, o que pode ser um limitador em diversos cenários.

Experiência dos avaliadores e Número de Avaliadores

O CCI agrupado encontrado para avaliadores experientes no índice intra-avaliador foi de 0.93 [0.88, 0.98], enquanto o encontrado para avaliadores inexperientes foi de 0.92 [0.87, 0.96]. Demonstrando uma pequena diferença entre avaliadores experientes e não experientes, algo que entra em acordo com o que dizem os criadores do FMS (Cuchna et al. (2016)). O fato de se tratar de uma confiabilidade em que o avaliador é comparado a ele mesmo em curtos períodos diferentes favorece esse valor, já que infere-se que o mesmo avaliador mantenha seu nível de conhecimento e capacidade de prognóstico ao longo de um curto período.

Sobre a confiabilidade inter-avaliador dos para avaliadores experientes, não foi possível a realização de uma meta-análise devido a existência de dois estudos que utilizaram apenas avaliadores experientes (Parenteau-g et al., 2014; Smith et al., 2017). Enquanto a meta-análise dos inexperientes indicou um CCI agrupado de 0.90 [0.83, 0.97], confiabilidade considerada excelente para os padres de Cicchetti(1994). Com relação à essa experiência, a coleta de dados apresentou uma grande diferença de experiência. Schneiders et al. (2011), Teyhen et al. (2012), Smith et al. (2013) e Morgan et al. (2023) treinaram avaliadores para a realização de seus estudos, Com tempo de treinamento variando de 4 horas com Morgan et al. (2023) a 20 horas com Teyhen et al. (2012) e Smith et al. (2013). Maeda et al. (2013). Shultz et al. (2013), Gulgin; Hoogenbomm (2014), Saki (2016) e Harper; Glass (2020) utilizaram avaliadores novatos, porém não mencionaram seu nível de treinamento. Em relação aos avaliadores experientes, Parenteau-g et al., (2014) e Smith et al., (2017) utilizaram avaliadores com certificação nível 1 no FMS, enquanto o restante dos artigos que utilizaram avaliadores experientes não mencionou a quantidade de treinamento de seus avaliadores.

Verificou-se a falta de padronização sobre o que os pesquisadores consideraram como experiente e novato. Sendo este um dos grandes problemas do estudos relacionados ao FMS. Essa falta de padronização compromete as evidências obtidas pela síntese estatística, já que o que um estudo considera como avaliador experiente pode ser considerado como um avaliador novato para outros artigos. Essa situação ocasiona a alocação de pessoas com diferentes períodos de uso do FMS no mesmo grupo de avaliadores. Lacunas também foram encontradas sobre as questões relacionadas ao tipo de treinamento que os sujeitos foram submetidos. Esses fatos dificultam a criação de evidências fortes sobre a influência da experiência dentro da confiabilidade do FMS. Além disso, é possível observar que um valor ruim de CCI encontrado provem de um grupo com pouca experiência na utilização do instrumento. No estudo de Gribble et al. (2013) um conjunto de estudantes encontraram um valor de CCI=0.372, o que é uma evidência que demonstra que a inexperiência pode ser capaz de afetar a confiabilidade do teste, apesar da já discutida questão do grande número de avaliadores utilizados nesse estudo.

Em contrapartida, Smith et al. (2013) demonstraram que a experiência dos avaliadores altera pouco os resultados de confiabilidade, a partir de 4 grupos com experiências diferentes e que obtiveram resultados semelhantes de CCI. Pode-se observar no estudo que além de não ser possível relacionar a experiência com valores maiores de confiabilidade, muitas vezes sua presença indica valores menores de CCIs. Isto pode ter sido ocasionado por uma visão superficial dos avaliadores novatos que, devido à menor



experiência, prendiam-se apenas aos detalhes mais evidentes dos movimentos e, assim, atingiram resultados mais parecidos quando comparados aos avaliadores experientes. De maneira geral, com a exceção do terceiro grupo de Gribble et al. (2013), que apresentou um nível ruim de confiabilidade, os estudos demonstraram resultados semelhantes entre as avaliações realizadas por experientes e inexperientes. Para ambos os tipos de avaliadores, a confiabilidade encontrada na maioria dos estudos (10 dos 13 incluídos) foi excelente, sendo que dois estudos foram bons e um estudo foi ruim, independentemente do nível de experiência que cada avaliador possuía com o FMS.

O número de avaliadores por estudo foi bem diversificado. No que diz respeito à confiabilidade intra-avaliador, a meta-análise demonstrou que, de maneira geral, poucos avaliadores tendem a apresentar uma confiabilidade menor, com exceção de alguns outliers. O quantitativo de avaliadores variou de um (Maeda et al., 2013; Shultz et al., 2013; Saki, 2016) a trinta e oito (Gribble et al., 2013). Dentro desses extremos temos os piores valores de confiabilidade intra-avaliador encontrados, com CCI=0.6 (Shultz et al., 2013) e com um CCI ruim de 0.372 (Gribble et al., 2013). Resultados como esses evidenciam que a análise a partir de apenas um avaliador para este índice pode não ser o ideal, já que o resultado fica suscetível ao desempenho de apenas um avaliador. Além disso, fica mais vulnerável à possíveis dificuldades do avaliador em pontuar algum dos sete testes, gerando uma grande variabilidade no valor total do FMS (Shultz et al., 2013). Dessa forma, parece que os índices de confiabilidade intra-avaliador são suscetíveis à grande variabilidade quando pautados, exclusivamente, em um avaliador. Além disso, grandes números de avaliadores também podem contribuir para uma maior variabilidade dos resultados dessa confiabilidade. O estudo que teve o maior CCI intra-avaliador foi o de Parenteau-g et al. (2014) com o cálculo realizado com base em dois avaliadores, indicando que este pode ser um número ideal para futuros estudos.

Em relação ao índice inter-avaliador, o número de avaliadores variou de dois (Maeda et al., 2013) a oito (Teyhen et al., 2012). A meta-análise demonstrou que uma menor quantidade de avaliadores indica uma melhor confiabilidade, vale destacar que o número de avaliadores apresentou pequena variabilidade, estando dentro do quantitativo de um a dez. Gribble et al., (2013) apresentaram três grupos diferentes de avaliadores que somados totalizaram trinta e oito. Esse número mais conciso, possivelmente contribuiu para números de CCI com menor variabilidade. Portanto, a menor variabilidade dos resultados para a confiabilidade inter-avaliador pode ter sido influenciada pelo menor número de avaliadores por estudo.

Número de Avaliados

As meta-análises para os índices intra-avaliador e inter-avaliador indicaram que quanto maior o número de avaliados, maior a confiabilidade do estudo. O número de avaliados disponíveis para avaliação em cada um dos estudos, tivemos um quantitativo de três (Gribble et al., 2013) a duzentos (Schneiders et al., 2011). Esses estudos também foram os que demonstraram o menor e o maior valor de CCI, respectivamente. De maneira geral, os valores de confiabilidade mais elevados foram encontrados nos estudos que utilizaram amostras maiores. Vale destacar que amostras homogêneas podem ser benéficas para potencializar os números de confiabilidade. Parenteau-g et al., (2014) e Smith et al., (2017) utilizaram amostras homogêneas de atletas em seus estudos e apresentaram excelentes índices de confiabilidade. Esse estudos apresentaram os CCIs mais elevados dessa revisão relativa ao índice inter-avaliador, com CCIs de 0.96 para Parenteau-g et al. (2014) e CCI de 0.88 para Smith et al. (2017) exemplificando como a homogeneidade de amostra pode ser benéfica para potencializar a confiabilidade.

Considerações metodológicas

Verificou-se que todos os artigos apresentaram baixa qualidade metodológica segundo os parâmetros do QAREL, com exceção do estudo de Teyhen et al. (2012) com qualidade total de 72,8%. O principal ponto de fragilidade dos estudos foi a qualidade interna, demonstrado nas perguntas de três a nove. Nenhum artigo apresentou percentual igual ou maior a 60 %. O que demonstra uma baixa qualidade dos estudos desta revisão sistemática. Dentro das perguntas que avaliam essa qualidade interna, a maioria se refere ao Blinding dos avaliadores a outros resultados, inclusive os seus, no caso da confiabilidade intra-avaliador (questão 4) e de outros avaliadores (questão 3), no caso do cálculo inter-avaliador.

De maneira geral, os estudos carecem de informações sobre esse Blinding e sobre o conhecimento ou não dos avaliadores sobre características clínicas (questão 6) e sobre características adicionais (questão 7). Essa falta de informações obriga um número de respostas como unclear, o que diminui a qualidade



metodológica dos artigos. Assim, tanto as amostras como os avaliadores, foram considerados como representativos para os objetivos dos estudos. Destaca-se como exceção os estudos que apresentaram resultado unclear para a amostra (Gribble et al., (2013); Maeda et al., 2013; Harper; Glass, 2020; Morgan et al., 2023) e que apresentaram resultado unclear para a representatividade dos avaliadores (Schneiders et al., 2011; Parenteau-G. et al., 2014; Smith et al., 2017).

Em relação a qualidade externa, demonstrado nas perguntas 1,2 e 10, todos os artigos apresentaram qualidade metodológica com percentuais que variaram de 66,7% a 100%. Todos os artigos utilizaram o teste correto, uma vez que o teste a ser aplicado é o próprio FMS para discutir a sua confiabilidade. Sobre a qualidade estatística, a grande maioria dos artigos apresentou resposta sim na única pergunta que avalia essa metodologia. Apenas Morgan et al., (2023) não apresentaram qualidade metodológica para a estatística, já que calculou o índice individual de confiabilidade pelo CCI e deveria ter usado o Kappa, uma vez que os valores dos testes individuais são variáveis ordinais.

Em suma, a qualidade dos artigos foi afetada principalmente pela carência de informações dos estudos. O que compromete as conclusões obtidas por essa revisão sistemática. O grande percentual de respostas como unclear dentro do checklist da QAREL indicam falta de informações principalmente nas questões relativas ao blinding dos avaliadores sobre os próprios resultados e sobre características dos participantes avaliados. Dessa forma, futuros estudos acerca do FMS devem proporcionar formas de ocultar esses resultados dos avaliadores, assim como demonstrar mais informações que possam servir para futuros revisões avaliarem a qualidade metodológica dos artigos.

Limitações do estudo

Os parâmetros do PRISMA foram utilizados para a condução dessa revisão sistemática, apesar disso, podemos não ter absorvido todos os estudos de confiabilidade do FMS. Isto ocorre porque limitou-se a periódicos revisados por pares e artigos publicados em língua inglesa, além de limitar as bases de dados investigadas. Apesar disso, selecionou-se algumas das principais bases relacionadas a intervenções clínicas e a análise de movimento, onde a grande maioria dos artigos disponibilizados estão na língua inglesa. Além disso, aplicou-se uma estratégia de busca ampla que recolheu muitos artigos. Dessa maneira, a possibilidade de algum artigo não absorvido mudar as conclusões dessa revisão é pequena.

Uma importante limitação está relacionada a grande heterogeneidade dos estudos, com altos valores de I^2 em grande parte das amostras, principalmente no que diz respeito a experiência dos avaliadores. Essa situação dificultou a realização da síntese estatística e da meta-análise, sendo necessário a criação de vários subgrupos. A dificuldade de alocação de um grande número de estudos em cada subgrupo também é uma limitação, já que para um poder estatístico consistente, são necessários pelo menos 5 estudos (Jackson;Turner, 2017). Isso não foi visto para os índices de; confiabilidade inter-avaliador com visualização em tempo real, confiabilidade inter-avaliador para visualização por vídeo e a confiabilidade inter-avaliador para avaliadores experientes, o que dificulta a inferência sobre a confiabilidade inter-avaliador de maneira geral. Além disso, estudos como Gulgin;Hoogenboon(2014) , Saki (2016) e Harper e Glass(2023) apresentaram falhas ao misturar avaliadores experientes e inexperientes para o cálculo do mesmo CCI, o que impediu estes estudos de entrarem na meta-análise do nível de experiência inter-avaliador.

Por último, nosso estudo teve como limitação na qualidade metodológica das fontes primárias, analisadas a partir do QAREL. No entanto, os artigos apresentaram boas características de validade externa e a estatística, indicando que as amostras de avaliados e avaliadores foram bem representativas e os cálculos de CCIs foram realizados da maneira correta. Entretanto, os estudos carecem de informações mais precisas relativas a como é feita a forma de visualização e o perfil total da amostra de avaliadores e avaliados, o que não é explícito na maioria dos artigos. Dessa forma, futuros estudos acerca da confiabilidade do FMS devem fornecer uma maior quantidade de informações para os revisores, de forma a evitar lacunas para a avaliação dos artigos. Além disso, futuros estudos devem procurar padronizar a classificação da experiência dos avaliadores indicando um corte que permita definir o que é um avaliador experiente ou não.

Conclusiones

Os resultados dessa revisão sistemática demonstram tendências que podem indicar o FMS como uma ferramenta de boa confiabilidade inter-avaliador e intra-avaliador, apesar das limitações metodológicas das fontes primárias. A experiência dos avaliadores parece não aumentar a confiabilidade do FMS, com a ressalva que a pior confiabilidade foi vista em um grupo de inexperientes, apesar do número de avaliadores também ter influenciado este valor. Em relação ao quantitativo de avaliadores, os melhores índices de confiabilidade foram encontrados quando dois avaliadores foram utilizados e grande número de avaliadores induzem a maior variabilidade dos resultados para confiabilidade. Além disso, foram encontradas pequenas diferenças para a confiabilidade quando a avaliação foi realizada através de vídeo, comparada à presencial, apesar da possibilidade de múltiplas visualizações quando o vídeo é utilizado. Sobre a influência do número de avaliados na confiabilidade do FMS, o fator preponderante é a homogeneidade da amostra e nem tanto o número de avaliados por estudo.

Tais aspectos permitem aos pesquisadores uma melhor compreensão de como realizar a aplicação do FMS, baseado na escolha dos avaliadores quanto a experiência e quantidade, de uma técnica de visualização que facilite a logística de aplicação do FMS nos avaliados, quantidade de avaliados e homogeneidade da amostra.

Por fim, destaca-se que essa revisão mostra evidências sobre as melhores formas de aplicação do FMS que pode ser usado nas práticas clínicas, esportivas e dentro de investigações científicas. Além disso, recomenda-se que estudos que recolham as fontes primárias de testes clínicos favoreçam a disseminação de práticas que estimulem sua reprodutibilidade e consistência de aplicação, o que aumenta a confiança e a precisão que os avaliadores têm sobre os resultados do teste, assim favorecendo sua disseminação e sua aplicação por diferentes indivíduos.

Agradecimientos

Agradecemos a colaboração do Laboratório de Treinamento de Força da UFJF (LABFOR), além do apoio financeiro para a publicação, em parte, pelo Programa de Pós-Graduação em Educação Física da Universidade Federal de Juiz de Fora (UFJF), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES) e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Financiamento

Os autores declaram que houve apoio financeiro para a pesquisa, autoria e/ou publicação deste artigo. Esta pesquisa foi financiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, com suporte ao pagamento da taxa de publicação (APC).

Referencias

- Asgari, M., Alizadeh, S., Sendt, A., & Jaitner, T. (2021). Evaluation of the Functional Movement Screen (FMS) in identifying active females who are prone to injury. A systematic review. *Sports Medicine-Open*, 7(1), 85. doi: 10.1186/s40798-021-00380-0.
- Bonazza, N. A., Smuin, D., Onks, C. A., Silvis, M. L., & Dhawan, A. (2017). Reliability, validity, and injury predictive value of the functional movement screen: a systematic review and meta-analysis. *The American journal of sports medicine*, 45(3), 725-732. doi: 10.1177/0363546516641937.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4), 284. doi: 10.1037/1040-3590.6.4.284.
- Cuchna, J. W., Hoch, M. C., & Hoch, J. M. (2016). The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. *Physical Therapy in Sport*, 19, 57-65. doi: 10.1016/j.ptsp.2015.12.002.



- Davies, K., Sacko, R. S., Lyons, M. A., & Duncan, M. J. (2022). Association between functional movement screen scores and athletic performance in adolescents: A systematic review. *Sports, 10*(3), 28. doi: 10.3390/sports10030028.
- dos Reis Molina, C., Borges Ribeiro Junior, D., da Silva Novaes, J., Costa dos Reis Souza, L., Veríssimo Perroux Lima, F., de Paula Nascimento, M. A., Zacaron Werneck, F., Figueiredo, A. J., Macedo Vianna, J., & Perroux de Lima, J. R. (2025). Prueba de coordinación motora KTK e iniciación deportiva en Brasil: una revisión sistemática. *Retos, 68*, 889-904. <https://doi.org/10.47197/retos.v68.112129>
- Gulgin, H., & Hoogenboom, B. (2014). The functional movement screening (FMS)TM: An inter-rater reliability study between raters of varied experience. *International journal of sports physical therapy, 9*(1), 14. doi: 10.1431/JSC.0b03dfjndu3e7
- Gould, Z. I., Oliver, J. L., Lloyd, R. S., Neil, R., & Bull, M. (2021). The Golf Movement Screen is related to spine control and X-factor of the golf swing in low handicap golfers. *The Journal of Strength & Conditioning Research, 35*(1), 240-246. doi: 10.1519/JSC.0000000000002664.
- Gribble, P. A., Brigle, J., Pietrosimone, B. G., Pfile, K. R., & Webster, K. A. (2013). Intrarater reliability of the functional movement screen. *The Journal of Strength & Conditioning Research, 27*(4), 978-981. doi: 10.1519/JSC.0b013e31825c32a8.
- Harper, B. A., & Glass, S. M. (2021). Item-level and composite-level interrater reliability of functional movement screenTM scores following condensed training in novice raters. *International Journal of Sports Physical Therapy, 16*(4), 1016. doi: 10.26603/001c.25793.
- Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports medicine, 30*(1), 1-15. doi: 10.2165/00007256-200030010-00001.
- Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods. 2017;8*(3):290-302. doi:10.1002/jrsm.1240
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj, 339*. doi: 10.1371/journal.pmed.1000100.
- Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of clinical epidemiology, 63*(8), 854-861. doi: 10.1016/j.jclinepi.2009.10.002
- Maeda, N., Urabe, Y., Fujii, E., Shinohara, H., Sasadai, J., Moriyama, N., ... & Yamamoto, T. (2013). The reliability of functional movement screen tm (fmstm) in the healthy young men. *13th AFSM Proceedings, 65.*, 65-68. doi: 10.2519/jospt.2012.3838.
- May, S., Chance-Larsen, K., Littlewood, C., Lomas, D., & Saad, M. (2010). Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. *Physiotherapy, 96*(3), 179-190. doi: 10.1016/j.physio.2009.12.002
- May, S., Littlewood, C., & Bishop, A. (2006). Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. *Australian journal of physiotherapy, 52*(2), 91-102. doi: 10.1016/s0004-9514(06)70044-7.
- McCunn, R., aus der Fünten, K., Fullagar, H. H., McKeown, I., & Meyer, T. (2016). Reliability and association with injury of movement screens: a critical review. *Sports medicine, 46*(6), 763-781. doi: 10.1007/s40279-015-0453-1.
- McCunn, R., aus der Fünten, K., Whalan, M., Sampson, J. A., & Meyer, T. (2018). Soccer injury movement screen (SIMs) composite score is not associated with injury among Semiprofessional soccer players. *journal of orthopaedic & sports physical therapy, 48*(8), 630-636. doi: 10.2519/jospt.2018.8037.
- Migliavaca CB, Stein C, Colpani V, Barker TH, Ziegelmann PK, Munn Z, Falavigna M; Prevalence Estimates Reviews-Systematic Review Methodology Group (PERSyst). Meta-analysis of prevalence: I² statistic and how to deal with heterogeneity. *Res Synth Methods. 2022 May;13*(3):363-367. doi: 10.1002/jrsm.1547. Epub 2022 Feb 23. PMID: 35088937.
- Moran, R. W., Schneiders, A. G., Mason, J., & Sullivan, S. J. (2017). Do Functional Movement Screen (FMS) composite scores predict subsequent injury? A systematic review with meta-analysis. *British journal of sports medicine, 51*(23), 1661-1669. doi: 10.1136/bjsports-2016-096938.

- Morgan, R., LeMire, S., Knoll, L., Schuster, E., Tietz, C., Weisz, A., & Schindler, G. (2023). The functional movement screen: exploring interrater reliability between raters in the updated version. *International Journal of Sports Physical Therapy*, 18(3), 737. doi: 10.26603/001c.74724.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, 5(1), 210. doi: 10.1186/s13643-016-0384-4.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372. doi: <https://doi.org/10.1136/bmj.n71>
- Parenteau-G, E., Gaudreault, N., Chambers, S., Boisvert, C., Grenier, A., Gagné, G., & Balg, F. (2014). Functional movement screen test: A reliable screening test for young elite ice hockey players. *Physical Therapy in Sport*, 15(3), 169-175. doi: 10.1016/j.ptsp.2013.10.001.
- Saki, F. (2016). Functional Movement Screen in elite boy basketball players: a reliability study. *Physical Treatments-Specific Physical Therapy Journal*, 6(4), 211-216. DOI: 10.18869/nrip.ptj.6.4.211
- Schneiders, A. G., Davidsson, Å., Hörman, E., & Sullivan, S. J. (2011). Functional movement screen™ normative values in a young, active population. *International journal of sports physical therapy*, 6(2), 75.
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *Bmj*, 349. doi: 10.1186/2046-4053-4-1.
- Shultz, R., Anderson, S. C., Matheson, G. O., Marcello, B., & Besier, T. (2013). Test-retest and interrater reliability of the functional movement screen. *Journal of athletic training*, 48(3), 331-336. doi: 10.4085/1062-6050-48.2.11.
- Silva, B., Rodrigues, L. P., Clemente, F. M., Cancela, J. M., & Bezerra, P. (2019). Association between motor competence and Functional Movement Screen scores. *PeerJ*, 7, e7270. doi: 10.4085/1062-6050-48.2.11.
- Smith, C. A., Chimera, N. J., Wright, N. J., & Warren, M. (2013). Interrater and intrarater reliability of the functional movement screen. *The Journal of Strength & Conditioning Research*, 27(4), 982-987. doi: 10.1519/JSC.0b013e3182606df2.
- Smith, P. D., & Hanlon, M. P. (2017). Assessing the effectiveness of the functional movement screen in predicting noncontact injury rates in soccer players. *The Journal of Strength & Conditioning Research*, 31(12), 3327-3332. doi: 10.1519/JSC.0000000000001757.
- Teyhen, D., Bergeron, M. F., Deuster, P., Baumgartner, N., Beutler, A. I., de la Motte, S. J., ... & O'Connor, F. (2014). Consortium for health and military performance and American College of Sports Medicine Summit: utility of functional movement assessment in identifying musculoskeletal injury risk. *Current sports medicine reports*, 13(1), 52-63. Doi: 10.2519/jospt.2012.3838.
- Santana, M., Salas Morillas, A., Peláez-Barrios, E., & López Bedoya, J. (2021). Calidad de movimiento en adolescentes practicantes y no practicantes de Gimnasia Acrobática mediante la batería Funtional Movement Screen. *Retos*, 41, 879-886. <https://doi.org/10.47197/retos.v41i0.86349>

Datos de los/as autores/as y traductor/a:

João Vitor de Souza Moreira	joao200865@hotmail.com	Autor/a
Matheus Oliveira Grandeni Pires	matheusopires@gmail.com	Autor/a
João Guilherme Vieira da Silva	joaoguilhermevds@gmail.com	Autor/a
Guilherme de Jesus do Nascimento	gnascimentoayla@gmail.com	Autor/a
Camille dos Reis Molina	camille.molina@hotmail.com	Autor/a e Traductor/a
Jeferson Macedo Vianna	jeferson.vianna@gmail.com	Autor/a
Jefferson da Silva Novaes	jeffsnovaes@gmail.com	Autor/a
José Vilaça-Alves	vilaca9@gmail.com	Autor/a
Aline Aparecida de Souza Ribeiro	ribeiroasaline@gmail.com	Autor/a
André Calil e Silva	andre_calil@hotmail.com	Autor/a

