

Desenvolvimento e Validação de Testes Físicos no âmbito militar: *Insights* para Atletas Táticos

Development and Validation of Physical Assessments in the military: insights for Tactical Athletes

RESUMO

As Forças Armadas militam constantemente para manter suas tropas em condições de agir quando demandado, o que pode ser expresso pelo conceito de prontidão para o combate. Neste sentido, torna-se essencial estruturar programas de treinamento físico apropriados e avaliações confiáveis, que juntos podem contribuir para o preparo adequado dos militares no desempenho de suas tarefas específicas. A literatura apresenta informações dispersas a respeito do desenvolvimento e validação de testes físicos para atletas táticos, tendo em vista que se trata de um processo em grande parte subjetivo, motivo pelo qual se faz necessário conhecer as melhores práticas e os requisitos fundamentais deste processo, a fim de se estruturar uma metodologia robusta. Portanto, o objetivo deste trabalho foi apresentar de forma sistematizada os principais *insights* para o desenvolvimento e validação de testes físicos para atletas táticos, além de apresentar a metodologia utilizada em uma validação que ocorreu no contexto da Força Aérea Brasileira, entre os anos de 2021 e 2023. Pesquisadores que pretendem desenvolver validações de testes físicos para atletas táticos, ou até mesmo empregadores de instituições civis cujo trabalho envolva altas demandas físicas, podem se valer destas informações como material de suporte para seus trabalhos.

Palavras-chave: Militares. Tarefas de Combate. Validação.

ABSTRACT

The Armed Forces constantly fight to keep their troops in a position to act when required, which can be expressed by the concept of combat readiness. In this sense, it is essential to structure good physical training programs and specific assessments, which together can contribute to the adequate preparedness of soldiers to perform their specific tasks. The literature presents scattered information regarding the development and validation of physical tests for tactical athletes, considering that it is a largely subjective process, which is why it is necessary to know the best practices and fundamental requirements of each stage, in order to structure a robust methodology. Therefore, the objective of this work was to systematically present the main insights for the development and validation of physical tests for tactical athletes, in addition to presenting a methodology used in a validation that took place among the Brazilian Air Force between the years of 2021 and 2023. Researchers who intend to develop validations of physical tests for tactical athletes, or even individuals from civilian institutions whose work involves high physical demands, can use this information as support material for their work.

Keywords: Combat Tasks. Military Personnel. Validation.

Willian Carrero Botta

Academia da Força Aérea – AFA, Pirassununga, SP, Brasil

Email: al_botta@hotmail.com

ORCID:

<https://orcid.org/0000-0002-2417-2814>

José Maurício Magraner

Universidade Estadual de Campinas – UNICAMP, Campinas, SP, Brasil

Email: jose_mauricio_z@hotmail.com

ORCID:

<https://orcid.org/0000-0001-6658-1221>

João Paulo Borin

Universidade Estadual de Campinas – UNICAMP, Campinas, SP, Brasil

Email: profjoaoborin@gmail.com

ORCID:

<https://orcid.org/0000-0002-7393-4053>

Received:	16 Jul 2024
Reviewed:	Jul/Nov 2024
Received after revised:	05 Dec 2024
Accepted:	13 Dec 2024



RAN

Revista Agulhas Negras

ISSN online 2595-1084

<http://www.ebrevistas.eb.mil.br/aman>



<https://creativecommons.org/licenses/by/4.0>



1 Introdução

Na última década, tem-se percebido um esforço por parte das Forças Armadas a nível mundial no sentido de direcionarem suas atenções a um conceito que é pétreo para os militares: a prontidão para o combate (Harrison, 2014). É sabido que a referida prontidão tem uma definição ampla, englobando fatores físicos, psicológicos e técnicos (atrelados ao ser humano), além de questões como capacidade logística e condições para realizar desdobramentos a qualquer tempo (Rounds, 2010). Porém, Nindl *et al.* (2012) defendem que o foco das atenções, quando se pensa em prontidão, precisa ser o fator humano, analisado de uma maneira holística.

Dentre todos os aspectos relacionados à prontidão, este artigo objetivou tratar do domínio físico, já que, independentemente das tecnologias que utilizam, os militares têm o seu corpo como instrumento de trabalho, do qual dependem a execução das Tarefas de Combate (TC) e, por consequência, o sucesso das operações (Nindl *et al.*, 2015). Moraes *et al.* (2008) afirmam que militares fisicamente bem preparados apresentam melhores níveis de prontidão quando comparados a sujeitos mau condicionados, além de se recuperarem mais rápido de lesões e apresentarem mais autoconfiança e motivação.

Nesse contexto, a literatura aponta para três vertentes que podem definir o domínio físico da prontidão: a teoria, a testagem e o treinamento. A primeira busca explicar cientificamente os fenômenos envolvidos no desempenho de TC, além de esclarecer quais são efetivamente os cenários esperados de atuação de determinada tropa. A segunda envolve as avaliações físicas propriamente ditas, cujo objetivo principal é identificar os sujeitos que estão aptos para executarem as TC de sua especialidade, além de identificar lacunas de capacidades físicas que precisam ser melhor desenvolvidas. Por fim, o treinamento busca alcançar o adequado preparo físico dos sujeitos, capacitando-os a desempenharem suas TC e, conseqüentemente, cumprirem suas missões com êxito, o que impacta diretamente os níveis de prontidão.

Quando o assunto é testagem, há uma quantidade considerável de informações disseminadas na literatura que buscam encontrar a metodologia adequada para desenvolver e validar avaliações confiáveis para atletas táticos, grupo que envolve os militares, agentes da segurança pública e demais profissionais que têm suas funções baseadas em demandas físicas, e que se diferem dos atletas convencionais na medida em que não seguem um calendário competitivo, não desempenham suas ações em condições ideais de repouso, alimentação, etc., além de apresentarem algumas peculiaridades, como a dedicação à proteção de seus pares, ao cumprimento dos deveres constitucionais e à preservação da vida (Alvar; Sell; Deuster, 2017).

Portanto, os objetivos deste trabalho foram i) apresentar de uma forma sistematizada os principais *insights* para o desenvolvimento e validação de testes físicos para atletas táticos e, a seguir,



ii) indicar uma metodologia utilizada em uma validação que ocorreu no contexto da Força Aérea Brasileira, entre os anos de 2021 e 2023, de forma que estes dados possam contribuir com organizações que estão conduzindo ou desejam conduzir este processo.

2 Referencial teórico

É importante destacar que, na busca por referenciais teóricos acerca do desenvolvimento de avaliações baseadas em demandas físicas de trabalho, não se pode focar somente nos estudos direcionados a populações compostas por atletas táticos, tendo em vista que outras populações também podem se beneficiar deste tipo de prática. Pode-se notar que diversos pesquisadores, militares ou não, têm considerado os conceitos estabelecidos por Constable e Palmer (2000) e por Payne e Harvey (2010), cujos artigos servem tanto para os atletas táticos como para setores e profissões em que são exigidas a execução de tarefas de trabalho com destacada demanda física. Estes autores apresentam três abordagens para a testagem física de tarefas de trabalho: i) testes preditores genéricos (assim como os testes físicos regulares de avaliação do condicionamento físico aplicados no âmbito das Forças Armadas do Brasil); ii) testes preditores baseados em tarefas (que almejam prever desempenhos específicos); iii) tarefas simuladas (abordagem que mais se aproxima da execução real de TC).

Payne e Harvey (2010) defendem que testes físicos que serão usados para selecionar pessoas para funções específicas precisam ter origem em uma criteriosa análise do trabalho, além de precisarem ser bons indicadores de tarefas essenciais, serem válidos (no sentido de medirem aquilo que se propõem a medir) e não serem discriminatórios. Estes autores também apresentam uma lista de 14 passos a serem considerados por aqueles que intentam operacionalizar testes focados em tarefas (TFT): i) identificação das tarefas; ii) seleção de tarefas a serem observadas; iii) observação das tarefas em campo; iv) análise das demandas físicas e riscos de lesão; v) identificação das tarefas mais críticas; vi) identificação dos padrões de desempenho; vii) desenvolvimento dos testes físicos; viii) análise da confiabilidade estatística; ix) coleta de dados da execução dos testes; x) estabelecimento de testes preditores; xi) identificação de padrões mínimos para os testes; xii) seleção da bateria final de testes; xiii) revisão dos parâmetros identificados para cada teste; xiv) ajustes de parâmetros em função de condições ambientais.

Robson *et al.* (2018) também desenvolveram pesquisas, junto à Força Aérea Americana, com objetivos similares, estabelecendo cinco etapas para a implantação de TFT, como segue: i) análise do trabalho e identificação das tarefas de demandas físicas; ii) identificação de testes válidos e confiáveis; iii) coleta de dados do desempenho dos sujeitos na execução das tarefas; iv) realização de cálculos estatísticos para avaliar a correlação dos testes com as tarefas; v) identificação dos



parâmetros mínimos das tarefas e dos testes desenvolvidos. Estes autores também apresentam dois caminhos possíveis a serem trilhados durante o processo de validação de TFT, sendo o primeiro deles a validação concorrente, onde uma amostra de sujeitos participa das avaliações físicas a fim de se medir a magnitude da correlação entre estes resultados e o desempenho em TC (ou desempenho no trabalho propriamente dito). E, o segundo caminho seria a validação preditiva, onde os resultados das avaliações são coletados por um determinado período de tempo (um ano por exemplo), para posteriormente os resultados serem comparados com as TC, investigando a capacidade de previsão de desempenhos entre as variáveis.

Milligan *et al.* (2016) contribuíram para esta área do conhecimento discorrendo sobre a relevância da validação e da análise da confiabilidade dentro do processo de desenvolvimento de TFT, o que influencia inclusive o estabelecimento dos pontos de corte (escores mínimos para aprovação). A validade guarda relação com a qualidade inferencial do que foi coletado, enquanto a confiabilidade trata-se de uma questão de consistência (reprodutibilidade) dos dados (Zumbo e Rupp, 2004), estando intimamente relacionadas. A determinação dos graus de confiabilidade das avaliações é, portanto, essencial, a fim de que os resultados não sejam amplamente influenciados pelos erros de medida (Payne; Harvey, 2010; Milligan *et al.*, 2016). Procedimentos estatísticos como correlações intraclasse e análises de variância estão entre os mais utilizados para investigar confiabilidade e efeito aprendido (Spiering *et al.*, 2012; Boyd *et al.*, 2015), trazendo maior clareza ao entendimento da variabilidade dos dados. Além disso, familiarizar previamente os sujeitos que serão testados com os protocolos e formas de execução das tarefas têm se apresentado como uma alternativa viável e efetiva para diminuir a variabilidade dos resultados, gerando mais estabilidade (Milligan *et al.*, 2016).

Tipton *et al.* (2013) apresentaram em seu estudo uma reflexão acerca dos critérios objetivos e subjetivos que precisam dialogar em um processo de desenvolvimento e validação de TFT, estruturando seis etapas fundamentais para tal: i) identificação de tarefas críticas por meio de análises do trabalho; ii) determinação das melhores práticas para a execução das tarefas; iii) concordância com relação aos parâmetros mínimos de desempenho das tarefas; iv) identificação das demandas físicas e fisiológicas da execução das tarefas; v) determinação da carga de trabalho relativa aceitável para a execução das tarefas; vi) de posse dos parâmetros mínimos, desenvolvimento das avaliações físicas. Os autores reforçaram, ainda, a importância de embasar ao máximo todas as decisões subjetivas que forem tomadas ao longo do processo, tendo em vista que a reprovação nas avaliações pode gerar problemas jurídicos, sendo necessário apresentar as fundamentações de tudo que foi estabelecido.

Outro aspecto essencial a ser apresentado é a participação de especialistas, que podem ser definidos como pessoas que executam ou supervisionam a execução de TC ou TS e que tenham experiência com isso (Blacklock *et al.*, 2015). Os especialistas podem ser muito úteis para determinar



as tarefas mais importantes de cada tropa, além de estabelecer quais os protocolos de execução destas tarefas e quais os parâmetros mínimos em termos de resultados (Milligan *et al.*, 2016). Blacklock *et al.* (2015) estabeleceram nove atributos desejáveis de um especialista, recomendando que pelo menos dois sejam atendidos, e este critério tem sido utilizado por diversos pesquisadores internacionais. Além disso, é interessante que os especialistas sejam representativos da população em questão, em termos de idade, sexo, experiência profissional e hierarquia. Existem diversas maneiras de proporcionar interação com os especialistas no decorrer da pesquisa, podendo-se citar como exemplos: entrevistas, reuniões em grupo, questionários, relatos de caso, exibição de filmagens das TC sendo executadas, entre outras. Recomenda-se que sejam usadas, ao menos, duas modalidades distintas de interação, e, de preferência, com especialistas diferentes.

Com relação aos pontos de corte, a literatura apresenta essencialmente duas abordagens para lidar com esta etapa: método baseado na análise da distribuição estatística dos dados (normativa), e método baseado no julgamento de especialistas, também conhecido como análise de critério (Cizek e Bunch, 2006). Os pontos de corte podem ser estabelecidos para cada teste individualmente, considerando que a reprovação em um deles determinará o fracasso do participante, ou pode ser composto por um valor global (escore final), permitindo que o participante compense os desempenhos ruins em algumas avaliações com desempenhos melhores em outras. Tudo vai depender do julgamento dos pesquisadores/especialistas no que diz respeito ao resultado final esperado da bateria de testes, aliado às expectativas depositadas nos avaliados. Além disso, os pontos de corte devem objetivar representar a performance mínima para obter sucesso em determinada tarefa, sendo exatamente o mesmo índice para todos, independente de sexo, idade e outras características descritivas, porque o foco é na tarefa a ser executada (que por definição será sempre a mesma independente de quem a executa).

Portanto, é possível notar que parece não existir uma única linha de ação para desenvolver e validar TFT, tratando-se de um processo em grande parte subjetivo (Tipton *et al.*, 2013) e assim verifica-se a necessidade de apresentar de forma sistematizada os principais insights para o desenvolvimento e validação de testes físicos para atletas táticos. As melhores práticas para a condução deste processo precisam ser pontuadas, visando objetivá-lo cada vez mais, o que pode contribuir para a consistência do respaldo legal da implantação de TFT.

3 Percurso Metodológico

Tendo resgatado os principais direcionamentos encontrados na literatura com relação ao desenvolvimento e validação de TFT, será apresentado a seguir o passo-a-passo que foi utilizado no contexto da Força Aérea Brasileira (FAB) focando em estabelecer Tarefas Simuladas para os cadetes



da Academia da Força Aérea, sintetizando a pesquisa realizada entre os anos de 2021 e 2023 (Botta, 2023). Em função de demandas específicas, a metodologia adotada foi resultado da união das melhores práticas executadas em outras pesquisas, além da tomada de decisões dos pesquisadores acerca de pontos que não estavam bem elucidados anteriormente.

Diante da tarefa de desenvolver e validar TFT para os cadetes da FAB, o primeiro passo foi realizar uma criteriosa análise do trabalho, tendo em vista que esta seria a base de tudo que viria a ser desenvolvido posteriormente. Foram reunidas e avaliadas todas as normas e legislações que regem o referido contexto profissional, de forma a entender o que os empregadores ou o próprio país esperam daqueles profissionais. Destas leituras, foram definidos cenários específicos de atuação, de onde futuramente seriam extraídas as primeiras versões das TC. Esta foi uma etapa da pesquisa de explorações iniciais, não havendo necessidade de envolver muitas pessoas, além de não ser necessário o engajamento obrigatório de especialistas no assunto.

A definição dos cenários de atuação precisou ser específica e bem delimitada, considerando que dela partiriam os mapeamentos das TC propriamente ditas. Além disso, neste momento os pesquisadores extrapolaram o raciocínio para além dos eventos que já haviam ocorrido ou até mesmo dos que eram bastante óbvios, buscando abarcar todas as possibilidades. Fatores como quantidade de pessoas envolvidas no cenário, ambiente, clima, equipamentos disponíveis, objetivo central das ações, tempo de duração, entre outros foram considerados.

A partir da definição dos cenários, passou-se a pensar na viabilidade de prosseguir no processo com todos os que haviam sido elencados. Cenários parecidos e que traziam demandas físicas similares foram agrupados, ou até mesmo um deles foi descartado. Não é possível determinar uma quantidade ideal de cenários a serem estabelecidos, tendo em vista que há muitas variáveis e peculiaridades atreladas a cada tipo de trabalho/missão, porém não se pode esquecer de que cada cenário será desmembrado em diversas TC, e um quantitativo elevado pode inviabilizar logisticamente o avanço da pesquisa.

O próximo passo consistiu em mapear as TC atreladas a cada cenário pré-estabelecido, sendo uma etapa desenvolvida somente pelos pesquisadores (já que estes tinham experiência profissional na área em questão). Nesta etapa também poderiam ser usados especialistas, para colaborar com os pesquisadores. As TC foram descritas de forma sucinta, sem muita riqueza de detalhes, pois o objetivo neste momento foi realizar um mapeamento amplo e abrangente. A principal meta desta etapa era gerar listas de TC vinculadas a cada um dos cenários pré-estabelecidos e, posteriormente, compará-las a fim de eliminar duplicidades, quando por exemplo a mesma TC fosse aplicável a mais de um cenário. Após este cruzamento de informações, foi possível estabelecer uma única lista de TC, sobre a qual se basearam as próximas etapas do processo.



Fez-se necessário, neste ponto, que a lista de TC elencada avançasse em seu processo de validação, que basicamente foi fundamentado na contribuição de especialistas. Pessoas que tinham vivência com as tarefas (executando-as ou observando-as) foram muito importantes neste momento. A estratégia de validação iniciou com a definição de três grupos de especialistas: i) G1: grupo formado por dez sujeitos que tinham comprovada experiência real na execução das TC; ii) G2: grupo formado por 71 sujeitos que tinham especialização operacional em busca e salvamento; iii) G3: grupo formado por quatro sujeitos com formação acadêmica em Educação Física e experiência com treinamentos e instruções militares.

O G1 interagiu com os pesquisadores por meio de entrevistas individuais semiestruturadas, com acesso à lista de TC previamente desenvolvida com antecedência e, após análises individuais, participaram das entrevistas. O foco destas entrevistas era avaliar a adequabilidade das tarefas e cenários elencados, permitindo que os especialistas fizessem sugestões de inclusão, exclusão ou alterações na descrição dos itens constantes no material. Em seguida, a lista de TC ajustada foi inserida em um formulário digital, que foi distribuído aos especialistas do G2. A intenção deste formulário era classificar as TC em graus de importância, valendo-se da escala proposta por Stein (2020), cujos escores variavam de um a cinco (1-não importante; 2-moderadamente importante; 3-muito importante; 4-missão essencial; 5-missão vital para sobreviver em combate). Outros autores sugerem que esta classificação englobe outros domínios além da importância, como frequência, duração e nível de exigência física (Robson *et al.*, 2018), mas a decisão pelos domínios que serão avaliados depende do julgamento dos pesquisadores, levando-se em consideração o teor de suas TC. Após a compilação dos dados dos formulários, foi avaliado se alguma TC deveria ser excluída da lista. O ponto de corte foi estabelecido na mediana 3 da distribuição, o que também pode ser adequado pelos pesquisadores em função de seus contextos de atuação.

Após inserir os ajustes em função da colaboração destes dois grupos de especialistas, pôde-se elaborar uma lista de final de TC. Nesse momento as TC foram descritas com maior riqueza de detalhes, já que até então elas haviam sido apresentadas de forma sucinta. O G3 começou a trabalhar neste momento, auxiliando os pesquisadores a elaborarem testes físicos baseados nas TC que haviam sido selecionadas. As interações ocorreram por meio de conversas em videoconferência, e o principal objetivo foi desenvolver avaliações que fossem representativas das TC, e que fossem viáveis de serem implementadas. Neste ponto, dois pesquisadores realizaram os testes propostos e os gravaram, a fim de detectar entraves logísticos ou impeditivos para a reprodução das avaliações em massa. Foram estabelecidas, no total, cinco TS, sendo dois testes realizados na água e o restante em terra firme.

A partir de então, os pesquisadores separaram uma amostra representativa de sujeitos oriundos da população para a qual os testes foram desenvolvidos, a fim de executarem as TS em caráter exploratório. Inicialmente foi planejada a realização das avaliações por três vezes consecutivas,



proporcionando os intervalos de descanso necessários para que os participantes pudessem se recuperar. A intenção dessas execuções foi a investigação de possível efeito-aprendizado (que poderia comprometer a confiabilidade do teste). Portanto, se houvesse necessidade deveriam realizar-se quantas execuções fossem necessárias até os testes estatísticos não apresentarem diferença significativa entre dois resultados sucessivos. O alcance de dois resultados consecutivos sem diferença estatística permitiu inferir a estabilidade do teste (sem necessidade de mais do que três execuções), e neste momento foram realizados os cálculos dos graus de confiabilidade, por meio do CCI. Quanto maior o grau de confiabilidade do teste melhor, porém os pesquisadores estabeleceram que o mínimo aceitável seria 75%.

Concluída a etapa de testes exploratórios, passou-se para a definição dos pontos de corte de cada avaliação, já que os testes não poderiam ser considerados completos se não houvesse uma previsão dos índices mínimos a serem alcançados. Importante destacar que quando uma avaliação aprova todos os sujeitos, isto não significa necessariamente que ela não está cumprindo sua função, porque é possível que apenas a existência do teste seja suficiente para motivar os sujeitos a treinarem (TIPTON *et al.*, 2013). Foi realizado um cálculo amostral e foram selecionados, aleatoriamente, 88 cadetes para participarem desta última etapa. Os sujeitos realizaram todas as TS uma única vez, tendo sido familiarizados com os testes previamente. De posse dos resultados, os pesquisadores passaram a analisar os dados unindo as abordagens normativa e de critério, a fim de valer-se dos pontos fortes de ambas. Foi estabelecido um ponto de corte inicial no percentil 85 da distribuição, e estes valores foram levados à discussão juntamente com o terceiro grupo de especialistas. A intenção foi determinar os pontos de corte para, longitudinalmente, avaliar a adequabilidade dos índices escolhidos, atualizando-os se necessário.

Finalmente, com os testes desenvolvidos, validados e de posse dos pontos de corte, o próximo passo (em construção) será aplicar as avaliações em maiores escalas/proporções, gerando dados para novas análises. Caso se perceba alguma lacuna nas avaliações, pode-se pensar em desenvolver novas TS, ou até mesmo em complementar a bateria de testagens com outros tipos de TFT, sempre visando a qualidade e a exequibilidade logística dos testes.

4 Considerações Finais

Este artigo teve por objetivo apresentar de uma forma sistematizada os principais *insights* para o desenvolvimento e validação de testes físicos para atletas táticos, tendo em vista que a literatura apresenta informações sobre esta temática de maneira dispersa, e que tarefas de demandas essencialmente físicas precisam ser desempenhadas por sujeitos capacitados, ou seja, “pessoa certa na função certa”. Pesquisadores que pretendem desenvolver validações de TFT, atletas táticos ou até



mesmo empregadores podem se valer destas informações como material de suporte para seus trabalhos. E, por se tratar de um processo que está em constante evolução e atualização, sugere-se que os interessados almejem tornar as etapas cada vez mais objetivas, além de aderir às melhores práticas que já se mostraram eficientes em outras validações.



Referências

- ALVAR, B.; SELL, K.; DEUSTER, P. Tactical strength and conditioning: an overview. **NSCA'S Essentials of Tactical Strength and Conditioning**. Champaign, IL: Human Kinetics, p. 1-8, 2017.
- BLACKLOCK, R. E. *et al.* Standard establishment through scenarios (SETS): a new technique for occupational fitness standards. **Work**, v. 52, n. 2, p. 375-383, 2015.
- BOTTA, Willian Carrero. **Modelo de Tarefas Simuladas para avaliação da prontidão para o combate no âmbito da Força Aérea Brasileira**. 2023. Tese (Doutorado em Educação Física) – Universidade de Campinas – UNICAMP, Campinas, 2023. Disponível em: <https://repositorio.unicamp.br/acervo/detalhe/1371250>. Acesso em: 12 dez. 2024.
- BOYD, L. *et al.* Variability in performance on a work simulation test of physical fitness for firefighters. **Applied Physiology, Nutrition, and Metabolism**, v. 40, n. 4, p. 364-370, 2015.
- CIZEK, G. J.; BUNCH, M. B. **Standard setting: A guide to establishing and evaluating performance standards on tests**. SAGE Publications Ltd, ed. 1, 2006.
- CONSTABLE, S.; PALMER, B. **The process of physical fitness standards development**. State of the Art Report. Human Systems Information Analysis Center, US Department of Defense, Washington, DC, 2000.
- HARRISON, Todd. Rethinking readiness. **Strategic Studies Quarterly**, v. 8, n. 3, p. 38-68, 2014.
- MILLIGAN, G. S. *et al.* Validity and reliability of physical employment standards. **Applied Physiology, Nutrition, and Metabolism**, v. 41, n. 6, p. S83-S91, 2016.
- MORAES, C. G. *et al.* A influência da frequência semanal do treinamento intervalado aeróbico, previsto no manual c20-20, sobre a potência aeróbica de militares. **Revista Brasileira de Prescrição e Fisiologia do Exercício**, São Paulo, v. 2, n. 8, 2008.
- NINDL, B. C. *et al.* Executive summary from the National strength and conditioning association's second blue ribbon panel on military physical readiness: military physical performance testing. **The Journal of Strength & Conditioning Research**, v. 29, p. S216-S220, 2015.
- NINDL, B. C. Strategies for Enhancing Military Physical Readiness in the 21st. 2012.
- PAYNE, W.; HARVEY, J. A framework for the design and development of physical employment tests and standards. **Ergonomics**, v. 53, n. 7, 2010.
- ROBSON, S. *et al.* Fit for duty? Evaluating the physical fitness requirements of battlefield airmen. **Rand health quarterly**, v. 7, n. 2, 2018.
- ROUNDS, M. The principal challenge of realizing total force fitness: Changing our readiness culture. **Military Medicine**, v. 175, n. suppl_8, p. 124-126, 2010.
- SPIERING, B. A. *et al.* Reliability of military-relevant tests designed to assess soldier readiness for occupational and combat-related duties. **Military medicine**, v. 177, n. 6, p. 663-668, 2012.
- STEIN, J. A. **Caffeine: Friend or foe in the modern combat environment?** 2020. PhD's Thesis (Doctor of Philosophy), Kansas State University, Manhattan, Kansas, 2020.
- TIPTON, M. J.; MILLIGAN, G. S.; REILLY, T. J. Physiological employment standards I. Occupational fitness standards: objectively subjective? **European journal of applied physiology**, v. 113, p. 2435-2446, 2013.
- Zumbo, B.D., Rupp, A.A. 2004. Responsible modeling of measurement data for appropriate inferences: important advances in reliability and validity theory. In *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. Edited by D. Kaplan. Sage Press, Thousand Oaks, Calif., USA. pp. 73–92.