



ISSN 2674-8169



Latindex



DOI



Questionários de avaliação da atividade física para gestantes: uma revisão da literatura

Carla Micheli da Silva ¹, Gleici da Silva Castro Perdoná ²



<https://doi.org/10.36557/2674-8169.2026v8n3p1885-1914>

Artigo recebido em 28 de Fevereiro e publicado em 31 de Março de 2026

ARTIGO DE REVISÃO

RESUMO

O objetivo deste estudo foi analisar a bibliografia referente à aplicação de questionários de atividade física em gestantes. Para tanto, foi realizada uma revisão da literatura hospedada em duas bases de dados, PubMed e Embase, por meio de seus respectivos mecanismos de busca. A critério, foram selecionados estudos com gestantes saudáveis, adultas, que mediram suas atividades físicas em diferentes contextos e que avaliaram pelo menos uma propriedade psicométrica (validade de conteúdo, critério, transculturalidade ou confiabilidade). Vinte e dois artigos foram considerados elegíveis, tendo sido a qualidade da evidência apresentada por cada artigo analisada com base em estudos de validade e confiabilidade para as três medidas: atividade física total, leve e moderada. Dos questionários identificados nesta revisão, se destacou o Questionário de Atividade Física na Gravidez (Pregnancy Physical Activity Questionnaire — PPAQ), por suas versões adaptadas para outros idiomas. Como resultado, o presente estudo explicitou que as propriedades psicométricas da maioria dos questionários analisados têm base em evidência de muito baixa à moderada, com exceção das versões portuguesa e inglesa para a medida de confiabilidade, que foram classificadas como evidência de alta qualidade. Em relação à validade de critério, os resultados foram insuficientes para as três medidas verificadas (física total, leve e moderada) e baseados em evidência de baixa qualidade. Os estudos de validade de conteúdo foram considerados adequados em relação a qualidade metodológica e os de validação transcultural, por fim, conseguiram cumprir todos os critérios estabelecidos por esse estudo.

Palavras-chave: atividade física, gestantes, questionários, revisão da literatura

Physical activity assessment questionnaire for pregnant women: a literature review

ABSTRACT

This study aims to analyze the literature concerning Physical Activity questionnaires used to assess physical activity among pregnant women. A literature review was conducted structured in two databases, PubMed and Embase, following their respective search guidelines. Studies conducted with healthy, adult pregnant women were considered, which measured their physical activities in different contexts and evaluated at least one psychometric property (content, criterion, transcultural, or reliability validation). Twenty-two articles were considered eligible, and the quality of evidence was analyzed in validity and reliability studies for three measures of vigorous, light, and moderate physical activity. Among the questionnaires identified in this review, the Pregnancy Physical Activity Questionnaire (PPAQ) prevailed, and its versions were adapted to other languages. The results of this study showed that the psychometric properties of most of the included questionnaires were based on evidence ranging from very low to moderate, except for the Portuguese and English versions for the reliability measurement, which were classified with high-quality evidence. Regarding criterion validation, the results were insufficient for the three measures assessed (vigorous, light, and moderate physical activity) and were based on low-quality evidence. Content validation studies were considered adequate in terms of methodological quality, and transcultural validation studies met all the criteria established by this study.

Keywords: physical activity, pregnant women, questionnaires, literature review, psychometric properties

Instituição afiliada – ¹ Pós-Graduação em Saúde Pública- Faculdade de Medicina de Ribeirão Preto -FMRP/USP ; ² Departamento de Medicina Social, Faculdade de Medicina de Ribeirão Preto -FMRP/USP

Autor correspondente: Gleici da Silva Castro Perdoná pgleici@fmrp.usp.br

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).





INTRODUÇÃO

A gestação é um período que demanda cuidados específicos com a saúde materno-fetal, entre eles a adoção de um estilo de vida saudável¹. Dentre os principais fatores que contribuem para a formação desse estilo de vida, a atividade física (AF) durante a gestação vêm sendo notavelmente enfatizada nos estudos elaborados nas últimas décadas, bem como a importância e os efeitos dessa prática regular²⁻¹⁰. Organizações Mundiais de saúde^{2,11,12} recomendam a prática de AF durante a gestação justamente por seus efeitos positivos na saúde da mulher e do bebê⁹, como: controle do peso gestacional, redução do diabetes gestacional (redução dos níveis de glicose e insulina materna e aumento da sensibilidade materna à insulina) e pré-eclâmpsia, bem como a redução no número de partos cesáreos^{2-4,6-8,10}. Além disso, a prática também influencia na qualidade do sono, no bem-estar e na autoestima da gestante^{3,7}, promovendo melhor qualidade de vida, saúde física e mental⁶.

Diante disso, se mostram fundamentais o monitoramento e a avaliação da AF durante a gravidez e, para tanto, se faz necessária a construção de instrumentos válidos e confiáveis. Nesse sentido, a aplicação de questionários se destaca como um instrumento relativamente acessível para avaliação dos níveis de AF habitual em estudos epidemiológicos, devido à facilidade de aplicação em grandes grupos e pelo menor custo envolvido^{8,10,13,14}. Esses questionários, no entanto, enquanto medida subjetiva, estão sujeitos a erros de autorrelato^{8,10,13,14}, pois ainda pesa a falta de conhecimento sobre como melhor quantificar a AF em mulheres grávidas com base nessa ferramenta¹³. Em contrapartida, as medidas objetivas como o acelerômetro, não apresentam o erro de autorrelato¹³, porém aspectos como alta custo e questão logística de aplicação tornam complexa a utilização deste tipo de medida em estudos epidemiológicos¹³.

A literatura analisada apresenta alguns questionários de AF, como o IPAQ¹⁵ (The International Physical Activity Questionnaire), o KPAS¹⁶ (Kaiser Physical Activity Survey), o GPAQ¹⁷ (Global Physical Activity Questionnaire), o AWAS¹⁸ (Australian Women's Activity Study), o PAPQ¹⁹ (Physical Activity and Pregnancy Questionnaire) e o PPAQ²⁰ (Pregnancy Physical Activity Questionnaire). Dos questionários citados, apenas dois (PAPQ e PPAQ) foram desenvolvidos especificamente para aplicação em gestantes. Assim, verifica-se que poucos questionários são adequados para a análise da condição específica da gravidez¹⁴. Além disso,



a pluralidade de questionários dificulta a comparabilidade dos níveis de AF entre estudos e países, especialmente quando há variação no emprego de medidas psicométricas²¹. Portanto, a seleção de um único questionário para análise da AF em gestantes é uma tarefa complexa, ainda mais pela escassez de questionários específicos para este demográfico e pela carência de estudos sobre as propriedades psicométricas desses questionários^{20,21}.

Uma revisão sistemática das propriedades psicométricas dos questionários de AF em gestantes e outras populações foi realizada por Sattler et al.²¹, em 2018. Não foi realizada, no entanto, uma avaliação da qualidade dos estudos no que tange a adaptação transcultural e tradução, especialmente do questionário PPAQ, principal ferramenta de pesquisadores no mundo todo¹⁴. Tampouco foram incluídos questionários de AF desenvolvidos ou adaptados para a população brasileira.

O objetivo desta revisão, portanto, foi especificamente identificar, resumir e comparar questionários de AF de gestantes com boas propriedades psicométricas. Foram consideradas, ainda, propostas de questionários para estudos de adaptação e validade transcultural.

METODOLOGIA

Para atingir os objetivos de trabalho, foi realizada uma revisão da literatura, na qual as questões delimitadoras foram: (1) “Existem questionários construídos especificamente para a população gestante que medem a AF nos contextos de domínio e intensidade?” (2) “Qual(is) apresenta(m) melhores avaliações em termos de medidas psicométricas?”. Para manter-se transparência na coleta, a ferramenta Prisma²² foi considerada e a revisão foi registrada no PROSPERO (CRD42023453279).

Os critérios de elegibilidade para levantamento dos artigos foram baseados nas recomendações feitas na revisão de 2018²¹, selecionando: i) estudos realizados com gestantes saudáveis e com idade superior a 16 anos; ii) questionários originalmente elaborados para mulheres grávidas; iii) estudos que tinham como objetivo a medição da AF, definida como qualquer movimento corporal produzido pelos músculos esqueléticos que resulta em gasto energético acima do nível de repouso²¹; iv) questionários que mediram pelo menos a frequência e/ou duração da AF; v) estudos que avaliaram a AF total e/ou as diferentes intensidades (leve, moderada e intensa) e contexto de AF (doméstica, ocupacional, locomoção e lazer); vi) estudos que avaliaram uma ou mais das seguintes propriedades psicométricas de



um questionário administrado: confiabilidade, validade de critério, validade de conteúdo e validade transcultural; vii) exclusivamente questionários autoadministrados ou por entrevista realizada pessoalmente, uma vez que diferentes tipos de coleta de dados possivelmente causariam heterogeneidade na coleta dos dados.

Por meio destes critérios, foram desconsiderados os trabalhos baseados em entrevistas por telefone ou outro meio (e-mail, por exemplo), bem como estudos em gestantes portadoras de alguma comorbidade, como: diabetes mellitus gestacional, hipertensão ou obesidade. Não se considerou limite quanto à idade máxima da população e nem quanto à semana gestacional das gestantes.

As fontes das informações foram as bases de dados eletrônicas *Public Medline* (PubMed) e Embase. A estratégia de busca incluiu variações dos termos “ATIVIDADE FÍSICA”, “QUESTIONÁRIO”, “GRAVIDEZ” e “GESTANTES”, provenientes dos Descritores em Ciência da Saúde (DeCS), e seus correspondentes na língua inglesa: “*physical activity*”, “*questionnaire*”, “*Pregnancy*” e “*pregnant*”, procedentes do *Medical Subject Headings* (MeSH). Em ambos os idiomas, recorreu-se aos operadores booleanos “AND” ou “OR”. (ver material suplementar para estratégia de busca completa - S1). A estratégia de busca foi adaptada para Embase e PubMed, seguindo suas diretrizes de busca individuais.

A primeira busca foi realizada em maio de 2019 e foi revisada em janeiro de 2023. Para a seleção dos artigos obtidos segundo os critérios de busca, dois pesquisadores foram eleitos (G e M) que, de forma independente, realizaram a seleção dos resumos e dos textos completos, bem como a extração de dados. Quando necessário, a discordância foi resolvida por um terceiro revisor.

Os estudos foram primeiramente agrupados por título, autor, ano de publicação, país de origem, título do periódico, objetivos, principais resultados e conclusões (Apêndice 1). Posteriormente, os resultados foram condensados em planilha do software Microsoft Office Excel®, constituindo-se, assim, o banco de dados dos estudos.

Para a extração dos dados, utilizou-se um formulário baseado na lista de verificação *Quality Assessment of Physical Activity Questionnaire (QAPAQ)*²³, no *Consensus Based Standards for the Selection of Health Measurement Instruments (COSMIN)*²⁴ e na revisão sistemática realizada por Sattler et al.²¹, em 2018.

A avaliação das propriedades de medidas (propriedades psicométricas: confiabilidade,



validade de conteúdo, adaptação transcultural e de critério) dos estudos selecionados foi realizada na seguinte sequência: primeiro, os resultados das propriedades psicométricas de cada estudo foram obtidos e avaliados. Posteriormente, a qualidade metodológica foi coletada e avaliada e, finalmente, a qualidade das evidências foi determinada. Para a avaliação dos resultados dos estudos e da qualidade metodológica foram seguidas as orientações do QAPAQ²³ e da revisão de 2018²¹. Dessa forma, a avaliação das propriedades psicométricas resultou na classificação dos estudos em: suficiente (+), insuficiente (-) e indeterminado (--).

Nos estudos de confiabilidade, considerou-se como suficientes os coeficientes de correlação intraclassa (ICC), Kappa e de concordância $\geq 0,70$ ou coeficientes de correlação de Pearson/Spearman $\geq 0,80$ ^{23,21}. Já para os estudos de validade de critério, as pontuações estão descritas no Suplemento 2 (VER S2). Considerando os estudos de validade de conteúdo e adaptação transcultural, optou-se por avaliar somente a qualidade metodológica dos estudos (VER QUADRO 1). Para a avaliação da qualidade metodológica dos estudos incluídos, foram atribuídos quatro níveis: Nível 1 (muito bom); Nível 2 (adequado), Nível 3 (duvidoso) e Nível 4 (inadequado). Essa avaliação foi aplicada individualmente nas medidas analisadas em cada estudo, por exemplo: AF Total, AF Moderada etc. As propriedades psicométricas foram escalonadas nos mesmos níveis, maiores detalhes podem ser obtidos no Suplemento 2.

Quanto à validade de critério, o tipo de instrumento de comparação e a semelhança dos construtos foram considerados. Caso o questionário foi avaliado com acelerômetro, níveis mais elevados (Nível 1 ou 2) foram atribuídos. Porém, se o instrumento de comparação foi outro questionário ou pedômetro, o nível atribuído foi rebaixado para Nível 3.

Considerando a semelhança do construto, níveis mais altas foram atribuídos quando, por exemplo, a AF moderada do questionário foi comparada com a AF moderada do acelerômetro (Nível 1) e níveis mais baixos (Nível 2 ou 3) foram atribuídos se o questionário foi comparado com medidas menos semelhantes ao construto, como questionários, pedômetros, ou diferentes níveis de intensidade.

Foi considerado adequado o artigo com o intervalo de tempo do teste e reteste de uma a duas semanas pois, dada a natureza da condição gestacional, é imprescindível considerar um intervalo curto, para garantir que a AF não se altere ao longo do tempo²¹.

No quesito qualidade de evidência, foram consideradas as propriedades psicométricas por questionário para três resultados: AF total, leve e moderada, separadamente, segundo a



abordagem GRADE (*Grading of Recommendation, Assessment, Development, and Evaluation*)²⁵. Desta forma, objetivou-se uma avaliação específica para cada questionário, independente da qualidade do estudo. Foram considerados quatro níveis de qualidade de evidência (alta, moderada, baixa ou muito baixa), baseado em quatro fatores^{21,24}: (1) Risco de vieses (qualidade metodológica): se os estudos, em sua maioria, foram classificados como Nível 1 de qualidade metodológica, e se foram fornecidas evidências de alta qualidade, os estudos não foram rebaixados. Caso ao contrário, se apresentaram estudos no Nível 3 ou 2, foram rebaixados em um nível. (2) Imprecisão: Rebaixou-se a qualidade de evidência em um nível se o tamanho da amostra não atendeu aos seguintes critérios: Para confiabilidade, $n = 45$ e para validade de critério, $n = 123$ ²¹. A qualidade da evidência foi rebaixada em dois níveis quando o tamanho amostral foi de $n < 12$ para confiabilidade ou $n < 32$ para validade de critério. (3) Inconsistência: rebaixou-se a qualidade de evidência em um ou dois níveis na presença de inexplicáveis diferenças nos resultados. (4) Indireta: os estudos que apresentaram evidências de populações diferentes da população de interesse tiveram a qualidade de evidência rebaixada em um ou dois níveis.

Nesta revisão, considerou-se a adaptação transcultural como critério para identificar se o artigo se tratava de uma proposta de questionário original ou de uma validação para outra língua/população, diferentemente do estudo de referência, Sattler et al²¹. Algumas etapas das metodologias deste artigo foram similares ao de Sattler et al.²¹ por própria recomendação na comparação dos estudos avaliados.

RESULTADOS

A busca bibliográfica resultou em 68 artigos, tendo sido selecionados 22 artigos para esta revisão (VER FIGURA 1 — Fluxograma). No geral, dos 22 estudos, 19 (86%)^{8,13,20,26,27-41,43} avaliaram a validade de conteúdo e/ou transcultural. A validade de critério foi avaliada por 11 (50%) artigos^{13,14,20,26,27,30,32,35,37,38,41} e 16 (73%) artigos^{8,13,14,20,26,27,30-32,34,35,37,38,40,41} avaliaram a confiabilidade (TABELAS 1,2,3). Do total de artigos (68), 31^{13,20,26-55} artigos eram duplicados; 11^{44-52,55,56} artigos foram excluídos com base na avaliação dos títulos e resumos (S1) e três⁵³⁻⁵⁵ foram excluídos com base na leitura completa por não terem avaliado nenhuma propriedade psicométrica.



Dos questionários identificados nesta revisão, o PPAQ^{14,20} prevaleceu com suas versões adaptadas para outros idiomas (20 artigos): turca^{13,38}, vietnamita²⁶, japonesa^{27,28}, brasileira²⁹, chinesa³⁰, polonesa^{8,31}, espanhola^{32,39}, dinamarquesa^{33,40}, árabe^{34,36}, sérvia³⁵, portuguesa³⁷, coreana⁴¹, persa⁴² e grega⁴³.

Na Tabela 1 foram resumidos os estudos de validação de conteúdo e transcultural. A maioria dos estudos utilizou métodos qualitativos para avaliar a validade de conteúdo e, conseqüentemente, alguns critérios foram modificados e/ou excluídos durante o processo de validação transcultural. O item mais removido das versões do PPAQ foi “cortar grama”, por não ser habitual nos países que receberam a tradução, e pequenas alterações semânticas foram necessárias no geral. Três estudos^{41,42,43} utilizaram métodos quantitativos e qualitativos para a validação de conteúdo, como: Razão de Validade de Conteúdo (RVC) ou Índice de Validade de Conteúdo (IVC). Na Versão inglesa²⁰ os autores fizeram um levantamento das AFs através do recordatórios de 24 horas para selecionar atividades relevantes para as gestantes. Os resultados do estudo mostraram que assistir televisão, andar a pé ou devagar no trabalho ao carregar cargas leves/moderadas e cuidados infantis foram as atividades relevantes. Somente a Versão polonesa³¹ recebeu nível três neste quesito, por não haver especificado o método para a avaliação do conteúdo. Os demais estudos utilizaram comumente o comitê de especialista e o estudo piloto para verificar a relevância dos conteúdos. Nesse estudo, a Versão polonesa³¹, considerou tanto o comitê quanto o estudo piloto para a validação de conteúdo, porém, também perpassando a validação transcultural. De acordo com os critérios utilizados para a avaliação da qualidade dos estudos de validação de conteúdo nenhum estudo alcançou o Nível 1 (muito bom), um dos fatores determinantes foi por a não utilização de métodos quantitativos. No caso da validação transcultural, todos os estudos fizeram uso de uma abordagem de tradução conhecida e atingiram o Nível 1 (muito bom).

Dos 11 estudos de validação, quatro deles^{20,27,30,37} usaram o acelerômetro como medida de comparação. Outros quatro empregaram outro questionário^{13,35,37,41} e dois^{13,26} usaram o pedômetro. Em relação a qualidade metodológica, três^{30,37,14} estudos alcançaram o Nível 1 e a maioria^{14,26,28,32,35,38,41} (sete estudos) receberam nível 3, como descrito na Tabela 1.

Considerando os resultados de confiabilidade, Tabela 2, todos os estudos atenderam os critérios do Nível 1. Desses, se destacaram as versões inglesa^{14,20}, vietnamita²⁶, turca^{14,38}, portuguesa³⁷, dinamarquesa⁴⁰, Sérvia³⁵ e Persa⁴⁵, por terem mostrado coeficientes de



correlação suficientes para todas as variáveis avaliadas. A Versão japonesa²⁷ do PPAQ mostrou confiabilidade suficiente para todos os escores, exceto para AF de transporte, esportes/exercícios e ocupacional, considerando o intervalo de uma semana. A Versão polonesa³¹ demonstrou resultado insuficiente para atividade leve e locomoção; a Versão árabe³⁴ mostrou insuficiência para as atividades vigorosas, transporte e ocupacional.

A Tabela 3 resumiu os resultados gerais da qualidade de evidência (ou seja, propriedades psicométricas suficientes/insuficientes) e qualidade de evidência (GRADE²⁵) para três desfechos de AF: AF total, leve e moderada (por questionário e por propriedade psicométrica). As versões inglesas^{14,20}, vietnamita²⁶, turcas^{13,28}, japonesa²⁷, chinesa³⁰, espanhola³², sérvia³⁵, portuguesa³⁷ e coreana⁴¹ (11 questionários) foram avaliadas quanto à confiabilidade e validade de critério. Após análise dos resultados das propriedades psicométricas (ou seja, confiabilidade e validade de critério), nenhum questionário foi suficiente em todas as medidas consideradas. A inconsistência só pôde ser avaliada para as versões que apresentaram dois questionários: as versões inglesas^{14,20} e as turcas^{13,28}. No geral, a qualidade de evidência dos questionários variou de muito baixa a alta.

Considerando as pontuações de confiabilidade, a maior parte dos estudos obtiveram resultados de moderada qualidade de evidência, com exceção da Versão sérvia³⁴ e coreana⁴¹ do PPAQ, que obtiveram a avaliação de qualidade muito baixa. As versões portuguesa³⁷ e inglesa¹⁴ mostraram confiabilidade suficiente ao avaliar a AF total, leve e moderada com base em evidências de alta qualidade

Nas pontuações de validade, os questionários obtiveram, em sua maioria, qualidade de evidência de muito baixa a moderada. A Versão inglesa²⁰ apresentou baixa qualidade para avaliar a AF total e leve e muito baixa qualidade para AF moderada. As versões vietnamita²⁶ e japonesa²⁷, obtiveram qualidade de evidência muito baixa para todas as pontuações e insuficiência para todas as propriedades psicométricas. As versões chinesa²⁷ e portuguesa³⁷ apresentaram qualidade moderada de evidência, porém insuficiência para todas as medidas. As outras versões obtiveram qualidade de evidência baixa^{13,28,32,35,41,14}. Não foi identificada nenhuma indireta grave e, portanto, não foi rebaixada a qualidade da evidência para nenhuma das pontuações da AF devido a esse fator. Apenas a Versão chinesa³⁰ e coreana⁴¹ mostraram propriedades psicométricas suficientes para avaliar a AF total, porém esses resultados foram baseados em evidência de qualidade muito baixa a baixa.



Considerando os resultados agrupados dos questionários e a confiabilidade, a maioria dos questionários demonstraram qualidade de evidência moderada ao avaliar a AF total. Nos resultados de validade, a maior parte dos estudos obtiveram qualidade de evidência baixa ao avaliar a AF total.

Considerando o conjunto de evidências das diferentes versões do PPAQ, houve alta qualidade de evidência, atribuindo ao PPAQ confiabilidade suficiente para avaliar a AF total, leve e moderada. Não se considerou rebaixar a qualidade de evidências para AF leve e moderada por causa das versões sérvia³⁵ e coreana⁴¹, que não avaliaram estas medidas. Os resultados de validade de critério do PPAQ foram inconsistentes para AF total (apenas 2 dois estudos foram suficientes), leve e moderada (três estudos foram suficientes). Os estudos suficientes, porém, foram baseados em evidência de qualidade baixa.

DISCUSSÃO

O objetivo dessa revisão foi verificar se haviam questionários com boas propriedades psicométricas medindo a AF na população gestante considerando os contextos e intensidade relevantes para esse demográfico. Este estudo então forneceu uma visão geral dos questionários de AF desenvolvidos especificamente para as gestantes, levando em consideração algumas propriedades psicométricas.

No processo de seleção de um questionário de avaliação de AF, é importante a clareza na descrição dos objetivos do questionário: o que pretende medir, para que finalidade e para qual população-alvo foi desenvolvido. É essencial, inclusive, a seleção de um questionário que tenha sido construído e validado para o público específico que pretende avaliar. Por exemplo, se o objetivo for medir a AF em adultos em geral, é adequada a aplicação do IPAQ¹⁵, ao mesmo passo que, se o objetivo for a medição da AF em mulheres com filhos pequenos, o AWAS¹⁸ deve ser considerado.

Esta revisão contribuiu com a avaliação da qualidade metodológica dos estudos em relação às propriedades psicométricas de validade de conteúdo e transcultural. A validade de conteúdo foi avaliada através do julgamento da relevância e abrangência dos itens⁵⁷, ou seja, se os itens do instrumento são relevantes para o construto a ser medido, para a população e para a finalidade, além de avaliar a aceitabilidade, compreensão e clareza dos itens⁵⁷. Apesar desse estudo eleger somente questionários que foram elaborados especificamente para o mesmo público (gestante), é importante que os estudos demonstrem como foi avaliada a



relevância e a abrangência dos itens. Isso por que, caso o instrumento seja posteriormente adaptado para outra população diferente da população-alvo original para a qual foi desenvolvido, deverá ser avaliado novamente se todos os itens são relevantes para esta nova população do estudo^{24,57}. Considerando os resultados dos estudos de validação do conteúdo, a maioria dos estudos forneceram alguma informação sobre como avaliaram a relevância dos conteúdos, porém, faltou um maior esclarecimento das etapas, como por exemplo, a verificação por parte das gestantes quanto à aceitabilidade e compreensão por elas dos itens. Na validação transcultural, todos os questionários utilizaram uma abordagem reconhecida de tradução e retro tradução, e a amostra do estudo foi semelhante à da população para o qual o questionário original foi desenvolvido, tendo recebido a avaliação “muito bom” (Nível 1) de acordo com os critérios deste estudo. Neste estudo não se avaliou os métodos estatísticos utilizados na validação transcultural, porém, a maioria dos estudos não utilizou esse tipo de método. Esta revisão forneceu uma visão geral da validade de conteúdo e transcultural dos questionários de AF para gestantes, entretanto, sabe-se que é necessária uma avaliação mais abrangente destas propriedades psicométricas.

No geral, é importante que os questionários desenvolvidos para gestantes deem maior atenção a estas propriedades psicométricas pois é essencial a seleção das atividades físicas relevantes durante a gestação²¹. Isso porque a gestação é um período em que a mulher está mais sensível às mudanças corporais, e a intensidade, duração, frequência e o tipo de AF podem mudar durante a gravidez²¹. Além disso, deve-se considerar o contexto em que a mulher vive. Por exemplo, o Brasil tem o número elevado de mulheres que tem como ocupação o trabalho doméstica. Segundo o Instituto Brasileiro de Pesquisa (IBGE)⁵⁸, em 2022, 1,5 milhão de mulheres ocupavam esta função. Assim, os questionários que avaliam a AF em mulheres brasileiras precisam se adaptar a essas especificidades do país. Essa questão, da ocupação doméstica, já foi levantada em outro estudo⁶⁰. O questionário adaptado para a população brasileira²⁹ apresentou qualidade adequada de conteúdo e muito boa de validação transcultural, porém, não abordou a questão da ocupação doméstica. Os autores, também, não apresentaram resultados quanto à validação ou confiabilidade dessa Versão.

No Brasil, existe o questionário de “Frequência de Atividade Física Diária” (QAFD) desenvolvido para gestantes, porém não atendeu os critérios de elegibilidade desta revisão. O QAFD foi desenvolvido por Takito et al.⁵⁹, e foi construído com base em uma lista de



atividades ligadas ao trabalho doméstica, ocupacional e de lazer realizados pelas gestantes brasileiras⁵⁹. Entretanto, de maneira semelhante ao PPAQ, as questões relacionadas à ocupação foram amplas/gerais e não consideraram as gestantes que tinham como ocupação o trabalho doméstica.

De acordo com a avaliação da validade de critério, a maioria dos estudos foram insuficientes, considerando a AF total e AF moderada. Em relação à qualidade metodológica dos estudos, somente as versões portuguesa²⁵ e inglesa³¹ atingiram o nível 1(muito bom), sendo que a maioria ficou entre a qualidade metodológica “adequada” (Nível 2) ou “duvidosa” (Nível 3), seja por não utilizarem o acelerômetro como instrumento comparativo ou por compararem diferentes intensidades entre o acelerômetro e o questionário. O acelerômetro (medida objetiva) e o questionário (medida subjetiva) foram os instrumentos de medida de comparação mais utilizados pelos estudos na avaliação da validade de critério.

Em relação aos estudos que analisaram a confiabilidade, todos foram considerados “muito bom” (nível 1) e demonstraram resultados suficientes para todas as pontuações avaliadas, exceto a Versão chinesa, que foi insuficiente para AF moderada. A maioria dos estudos utilizou CCI e intervalos de tempo adequados entre teste e reteste.

Os resultados desta revisão mostraram que a qualidade da evidência dos questionários avaliados é alta para avaliar a propriedade psicométrica de confiabilidade para as AF total, leve e moderada, enquanto a qualidade de evidência para os estudos de validade de critério é baixa. Nenhum questionário demonstrou suficiência para todos os escores de AF medidos. Agrupando todas as versões do PPAQ, considerou-se neste estudo a qualidade de evidência de “baixa” a “moderada”, corroborando com estudo de Sattler et al.²¹. Apesar desse estudo ter apontado questionários que alcançaram o Nível 1 na qualidade metodológica, tanto nos estudos de validação como de confiabilidade, na qualidade de evidência esses questionários não atingiram alta qualidade para essas propriedades psicométricas. Uma limitação desse estudo está na relação entre a validação e a variedade de instrumentos utilizados para a comparação, o uso de diferentes pontos de cortes e diferentes tipos de processamentos de dados. Todos esses fatores fazem com que a qualidade metodológica destes estudos varie consideravelmente.

Devido a essas deficiências, estudos futuros de validação devem relatar detalhadamente suas regras de decisão e devem desenvolver diretrizes para o uso ideal dos dados do



acelerômetro na população-alvo (por exemplo, gravidez). Nesse estudo não foi avaliada a capacidade de resposta nem o erro padrão, o que pode ser uma limitação. Entretanto, na revisão de Sattler et al., poucos estudos avaliaram essas medidas.

Em geral, recomenda-se a utilização de questionários que foram avaliados na população de interesse e que forneceram resultados consistentes quanto às propriedades psicométricas (validade de conteúdo, validade de critério e confiabilidade) com base em evidências de alta qualidade. É importante que os estudos forneçam mais informações sobre as abordagens que foram utilizadas na validação de conteúdo, como também, na avaliação das outras propriedades psicométricas. Recomenda-se que estudos de construção ou adaptação de questionários de AF para gestantes brasileiras considerem a ocupação doméstica no contexto ocupacional, já que é uma profissão relevante na população feminina do país. Recomenda-se, também, que estudos futuros de validação desenvolvam diretrizes para o processamento e uso do acelerômetro em gestantes para facilitar a comparabilidade entre os estudos e, também, que desenvolvam pontos de corte específicos para a população gestante.

Corroborando o estudo de Sattler et al.²¹, o PPAQ e suas versões ainda representa a melhor escolha para avaliação da AF durante a gestação. Entretanto, algumas versões do PPAQ mostraram propriedades psicométricas insuficientes para algumas medidas. Versões recentes, no entanto, (como a portuguesa e inglesa) mostraram um aumento nas evidências para avaliar a confiabilidade nas medidas de AF total, leve e moderada. Há um movimento no meio científico, portanto, em direção à melhoria da qualidade de evidências dos questionários de AF para gestantes e as revisões de literatura, como a de Sattler et al.²¹, contribuem para isso. De forma geral, esta revisão conseguiu responder as questões delimitadoras (descritas nos métodos) que são fundamentais para os estudos de revisão como orientadas por estudos⁶¹.

CONSIDERAÇÕES FINAIS

O PPAQ e suas versões se estabeleceram enquanto os questionários mais indicados até o momento para avaliação da AF em gestantes. Entretanto, os resultados deste estudo mostraram que as propriedades psicométricas da maioria dos questionários incluídos foram baseadas em evidência de muito baixa a moderada, com exceção das versões portuguesa³⁷ e inglesa¹⁴ para a medida de confiabilidade. Porém, ao considerar os resultados agrupados das



diferentes versões do PPAQ, houve evidência de alta qualidade e confiabilidade suficiente na avaliação da AF total leve e moderada. Em relação a validade de critério, os resultados foram insuficientes para as três medidas e baseados em evidência de baixa qualidade. Os estudos de validade de conteúdo foram considerados adequados em relação à qualidade metodológica e os estudos de validação transcultural alcançaram o Nível 1 pelos critérios estabelecidos por este estudo.

REFERÊNCIAS

1. Freitas IGCD, Lima CDA, Santos VM, Silva FT, Rocha JSB, Dias OV. Nível de atividade física e fatores associados entre gestantes: estudo epidemiológico de base populacional. *Cien Saúde Colet.* 2022;27(11):4315-28.
2. Physical Activity and Exercise During Pregnancy and the Postpartum Period: ACOG Committee Opinion, Number 804. *Obstet Gynecol.* 2020;135(4):e178-88.
3. Gjestland K, Bø K, Owe KM, Eberhard-Gran M. Do pregnant women follow exercise guidelines? Prevalence data among 3482 women, and prediction of low-back pain, pelvic girdle pain and depression. *Br J Sports Med.* 2013;47(8):515-20.
4. Harrod CS, Chasan-Taber L, Reynolds RM, Fingerlin TE, Glueck DH, Brinton JT, et al. Physical activity in pregnancy and neonatal body composition: the Healthy Start study. *Obstetrics and gynecology.* 2014; 124(201): 257.
5. Ko Y L, Chen, CP, Lin PC. Physical activities during pregnancy and type of delivery in nulliparae. *European journal of sport Science.*2016; 16(3):374-80.
6. Mourady D, Richa S, Karam R, Papazian T, Hajj Moussa F, El Osta N, et al. Associations between quality of life, physical activity, worry, depression and insomnia: A cross-sectional designed study in healthy pregnant women. *PloS one.* 2017; 12(5), e0178181.
7. Smith KM, Campbell CG. Physical activity during pregnancy: Impact of applying different physical activity guidelines. *Journal of Pregnancy,* 2013;165617.
8. Suliga E, Sobaś K, Król G. Validation of the Pregnancy Physical Activity Questionnaire (PPAQ). *Med Stud.* 2017; 33(1): 40-45
9. Barakat R, Refoyo I, Coteron J, Franco E. Exercise during pregnancy has a preventative effect on excessive maternal weight gain and gestational diabetes. A randomized controlled trial. *Braz J Phys Ther.* 2019;23(2):148-55.



10. Santos PC, Maciel LYS, Abreu S, Mesquita AR, Mesquita CC, Lopes S, et al. Cultural adaptation and validation of the "Pregnancy Physical Activity Questionnaire" for the Portuguese population. *PLoS One*. 2023; 18(1):e0279124.
11. World Health Organization, T. (2010). *Global recommendations on physical activity for health*. World Health Organization.
12. ACSM AC. SM. Impact of Physical Activity during Pregnancy and Postpartum on Chronic Disease Risk. *Medicine & Science in Sports & Exercise*. 2006;38(5), 989–1006
13. Çırak Y, Yılmaz GD, Demir YP, Dalkılıç M, Yaman S. Pregnancy physical activity questionnaire (PPAQ): reliability and validity of Turkish version. *J Phys Ther Sci*.2015; 27(12): 3703-09.
14. Chasan-Taber L, Park S, Marcotte RT, Staudenmayer J, Strath S, Freedson P. Update and Novel Validation of a Pregnancy Physical Activity Questionnaire. *Am J Epidemiol*. 2023;192(10):1743-53.
15. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & science in sports & exercise*.2003;35(8), 1381-95.
16. Schmidt MD, Freedson PS, Pekow P, Roberts D, Sternfeld B, Chasan-Taber L. Validação do Kaiser Physical Activity Survey em mulheres grávidas. *Medicina e Ciência no Esporte e Exercício*. 2006;38(1), 42–50.
17. Armstrong T, Bull F. Development of the world health organization global physical activity questionnaire (GPAQ). *Journal of Public Health*.2006; 14, 66-70.
18. Fjeldsoe BS, Marshall AL, Miller YD. Measurement properties of the Australian Women's Activity Survey. *Med Sci Sports Exerc*. 2009;41(5):1020-33.
19. Haakstad LA, Voldner N, Henriksen T, Bø K. Physical activity level and weight gain in a cohort of pregnant Norwegian women. *Acta Obstet Gynecol Scand*. 2007;86(5):559-64.
20. Chasan-Taber L, Schmidt MD, Roberts DE, Hosmer D, Markenson G, Freedson PS. Development and validation of a Pregnancy Physical Activity Questionnaire. *Med Sci Sports Exerc*. 2004;36(10):1750-60.
21. Sattler MC, Jaunig J, Watson ED, van Poppel MNM, Mokkink LB, Terwee CB, et al. Physical Activity Questionnaires for Pregnancy: A Systematic Review of Measurement Properties. *Sports Med*. 2018;48(10):2317-46.
22. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Int J Surg*. 2021;88:105906.



23. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. *Sports Med.* 2010;40(7):525-37.
24. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. Protocol of the COSMIN study: CONsensus-based Standards for the selection of health Measurement Instruments. *BMC Med Res Methodol.* 2006;6(2):1-7.
25. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol.* 2011;64(4):380-2.
26. Ota E, Haruna M, Yanai H, Suzuki M, Anh DD, Matsuzaki M, et al. Reliability and validity of the Vietnamese version of the Pregnancy Physical Activity Questionnaire (PPAQ). *Southeast Asian J Trop Med Public Health.* 2008;39(3):562-70.
27. Matsuzaki M, Haruna M, Nakayama K, Shiraishi M, Ota E, Murayama R, et al. Adapting the Pregnancy Physical Activity Questionnaire for Japanese pregnant women. *J Obstet Gynecol Neonatal Nurs.* 2014;43(1):107-16.
28. Matsuzaki M, Haruna M, Ota E, Yeo S, Murayama R, Murashima S. Translation and cross-cultural adaptation of the Pregnancy Physical Activity Questionnaire (PPAQ) to Japanese. *Biosci Trends.* 2010; 4(4):170-7.
29. Silva FT, Araujo Júnior E, Santana EF, Lima JW, Cecchino GN, Silva Costa FD. Translation and cross-cultural adaptation of the Pregnancy Physical Activity Questionnaire (PPAQ) to the Brazilian population. *Ceska Gynekol.* 2015; 80(4):290-8.
30. Xiang M, Konishi M, Hu H, Takahashi M, Fan W, Nishimaki M, et al. Reliability and Validity of a Chinese-Translated Version of a Pregnancy Physical Activity Questionnaire. *Matern Child Health J.* 2016; 20(9):1940-7.
31. Krzepota J, Sadowska D, Sempolska K, Pelczar M. Measuring physical activity during pregnancy - Cultural adaptation of the Pregnancy Physical Activity Questionnaire (PPAQ) and assessment of its reliability in Polish conditions. *Ann Agric Environ Med.* 2017; 24(4): 640-43.
32. Oviedo-Caro MÁ, Bueno-Antequera J, Munguía-Izquierdo D. Transcultural adaptation and psychometric properties of Spanish version of Pregnancy Physical Activity Questionnaire: the PregnActive project. *Gac Sanit.* 2019;33(4):369-76.
33. Schmidt JT, Nielsen J, Riis A, Larsen BT. Translation and cultural adaptation of the Pregnancy Physical Activity Questionnaire into Danish using the dual-panel approach: comparison with outcomes from an alternative translation approach. *BMC Res Notes.* 2021;14(1):225.
34. Younis AS, Alyousefi NA, Al-Habib DM, Al-Omran AT. The validation and cross-cultural

- adaptation of the Arabic version of the pregnancy physical activity questionnaire. *Saudi Med J*. 2021;42(5): 499.
35. Rovcanin M, Jankovic S, Mikovic Z, Sipetic Grujicic S, Ersk IRB, Lackovic M, et al. The Translation and Cross-Cultural Adaptation of the Pregnancy Physical Activity Questionnaire: Validity and Reliability of a Serbian Version (PPAQ-SRB). *Healthcare (Basel)*. 2022;10(8):1482.
36. Papazian T, El Osta N, Hout H, Chammas DE, El Helou N, Younes H, et al. Pregnancy physical activity questionnaire (PPAQ): Translation and cross-cultural adaption of an Arabic version. *PLoS One*. 2020;15(3):e0230420.
37. Santos PC, Maciel LYS, Abreu S, Mesquita AR, Mesquita CC, Lopes S, et al. Cultural adaptation and validation of the "Pregnancy Physical Activity Questionnaire" for the Portuguese population. *PLoS One*. 2023; 18(1): e0279124.
38. Tosun OC, Solmaz U, Ekin A, Tosun G, Mutlu EK, Okyay E, et al. The Turkish version of the pregnancy physical activity questionnaire: cross-cultural adaptation, reliability, and validity. *Rev Pesqui Fisioter*. 2015; 27(10): 3215-21.
39. Coll-Risco I, Camiletti-Moirón D, Acosta-Manzano P, Aparicio VA. Translation and cross-cultural adaptation of the Pregnancy Physical Activity Questionnaire (PPAQ) into Spanish. *J Matern Fetal Neonatal Med*. 2019;32(23):3954-61.
40. Krøner FH, Knudsen SP, Roland CB, Alomairah SA, Molsted S. Validity and reliability of the Danish version of the pregnancy physical activity questionnaire to assess levels of physical activity during pregnancy. *J Matern Fetal Neonatal Med*. 2022;35(23):4566-72.
41. Han JW, Kang JS, Lee H. Validity and reliability of the Korean version of the pregnancy physical activity questionnaire. *J. Environ. Health*. 2020; 17(16): 5873.
42. Kazemi AF, Hajian S, Sharifi N. The psychometric properties of the Persian version of the pregnancy physical activity questionnaire. *Int J Women's Health Reprod*. 2019; 7(1): 54-60.
43. Mitrogiannis I, Chatzakis C, Skentou C, Koutalia N, Makrydimas S, Efthymiou A, et al. Pregnancy physical activity questionnaire: Translation and cross-cultural adaptation of a Greek version. *Eur J Obstet Gynecol Reprod Biol*. 2023;291:156-61.
44. Chandonnet N, Saey D, Alméras N, Marc I. French Pregnancy Physical Activity Questionnaire compared with an accelerometer cut point to classify physical activity among pregnant obese women. *PLoS One*. 2012;7(6):e38818.
45. Bernardo D, Carvalho C, Leirós-Rodríguez R, Mota J, Santos PC. Comparison of the Portuguese Version of the Pregnancy Physical Activity Questionnaire (PPAQ) with Accelerometry for Classifying Physical Activity among Pregnant Women with Obesity. *Int J Environ Res Public*



- Health. 2023;20(2):929.
46. Adanaş Aydın G, Taşan HA, Tarhan N, Çakar E, Şenol Güler N, Ankaralı H, et al. Reliability and validity of Turkish version of pregnancy physical activity questionnaire (PPAQ) in patients with gestational diabetes mellitus. *J Obstet Gynaecol.* 2020;40(2):176-81.
47. Cohen TR, Plourde H, Koski KG. Use of the Pregnancy Physical Activity Questionnaire (PPAQ) to identify behaviours associated with appropriate gestational weight gain during pregnancy. *J Phys Act Health.* 2013;10(7):1000-7.
48. Watson ED, Micklesfield LK, van Poppel MNM, Norris SA, Sattler MC, Dietz P. Validity and responsiveness of the Global Physical Activity Questionnaire (GPAQ) in assessing physical activity during pregnancy. *PLoS One.* 2017;12(5):e0177996.
49. Aittasalo M, Pasanen M, Fogelholm M, Ojala K. Validity and repeatability of a short pregnancy leisure time physical activity questionnaire. *J Phys Act Health.* 2010;7(1):109-18.
50. Chasan-Taber L, Schmidt MD. Pregnancy Physical Activity Questionnaire. *Can J Public Health.* 2016;106(8):e563.
51. Wojtyła C, Ciebiera M, Wojtyła-Buciora P, Janaszczyk A, Brzęcka P, Wojtyła A. Physical activity patterns in third trimester of pregnancy - use of pregnancy physical activity questionnaire in Poland. *Ann Agric Environ Med.* 2020;27(3):388-93.
52. Brett KE, Wilson S, Ferraro ZM, Adamo KB. Self-report Pregnancy Physical Activity Questionnaire overestimates physical activity. *Can J Public Health.* 2015;106(5):e297-302.
53. Siyad S, Mustafa N, Banu SS, Buharoon MA, Al Mulla MA, Ankarali H, Carrick FR, Abdulrahman M. Pregnant women's perceptions of physical activity: Adaptation of the pregnancy physical activity questionnaire in Dubai, a multicultural society. *J Public Health Res.* 2021;11(1):2261.
54. Antosiak-Cyrak KZ, Demuth A. A study of physical activity levels of pregnant women using the Polish version of Pregnancy Physical Activity Questionnaire (PPAQ-PI). *Ginekol Pol.* 2019;90(5):250-55.
55. Conway MR, Marshall MR, Schlaff RA, Talge NM, Pfeiffer KA, Pivarnik JM. Validity of the pregnancy physical activity questionnaire for maternal recall. *Measurement in Physical Education and Exercise Science.* 2020;24(4), 264-72.
56. Chandonnet N, Saey D, Dufresne A, Robitaille J, Marc I. Assessment of physical activity in pregnant obese women using the pregnancy physical activity questionnaire (PPAQ) and accelerometer. *Canadian Journal of Diabetes.* 2011;35(2), 148.
57. Cunha CM, Almeida Neto OP, Stackfleth R. Principais métodos de avaliação psicométrica da validade de instrumentos de medida. *Revista de Atenção à Saúde.* 2016;14(47), 75-83.



58. IBGE 2022
59. Takito MY, Neri LDCL, Benício MHA. Avaliação da reprodutibilidade e validade de questionário de atividade física para gestantes. *Revista Brasileira de Medicina do Esporte*. 2008;14(2), 132-38.
60. Silva CM, Sartorelli DS, Machado DRL, Perdoná GDSC. Padrão de atividade física em gestantes usuárias do Sistema Público de Saúde. *Conexões*. 2021; 19, e021003-e021003.
61. Galvão MCB, Ricarte ILM. Revisão sistemática da literatura: conceituação, produção e publicação. *Logeion: Filosofia da informação*. 2019; 6(1), 57-73.
62. Saglam M, Arikan H, Savci S, Inal-Ince D, Bosnak-Guclu M, Karabulut E, et al. International physical activity questionnaire: reliability and validity of the Turkish version. *Percept Mot Skills*. 2010 ;111(1):278-84.
63. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S442-9.
64. Swartz AM, Strath SJ, Bassett DR Jr, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S450-6.
65. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777-81.
66. Schneider PL, Crouter S, Bassett DR. Pedometer measures of free-living physical activity: comparison of 13 models. *Med Sci Sports Exerc*. 2004;36(2):331-5.
67. Kumahara H, Schutz Y, Ayabe M, Yoshioka M, Yoshitake Y, Shindo M, et al. The use of uniaxial accelerometry for the assessment of physical-activity-related energy expenditure: a validation study against whole-body indirect calorimetry. *Br J Nutr*. 2004;91(2):235-43.
68. Hildebrand M, Van Hees VT, Hansen BH, et al. Age group comparability of raw accelerometer output from wrist-and hip-worn monitors. *Med Sci Sports Exerc*. 2014;46(9): 1816-24.
69. Montoye AHK, Clevenger KA, Pfeiffer KA, et al. Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. *J Sports Sci*. 2020;38(22):2569-78.

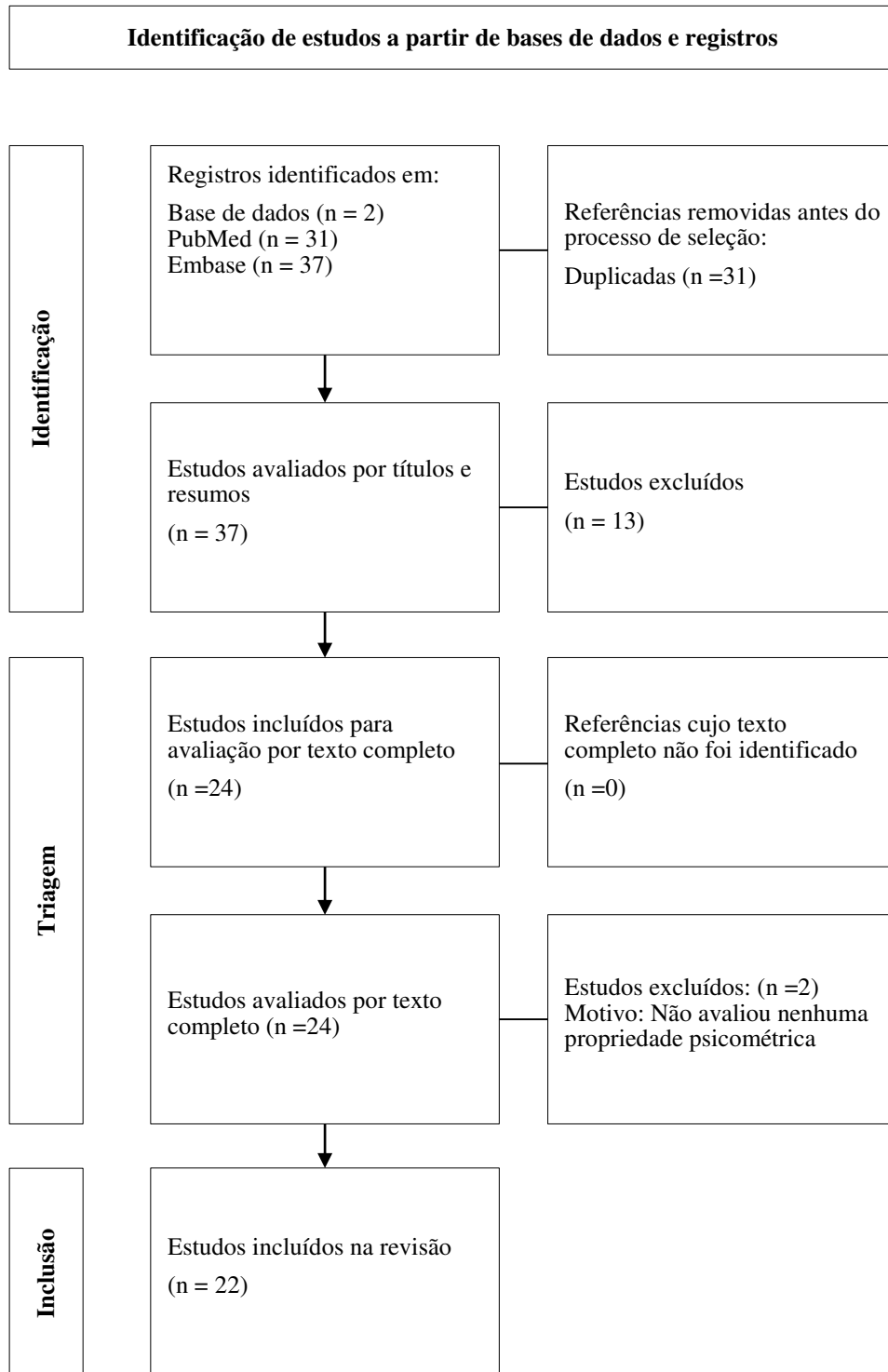


FIGURA 1 – Fluxograma da seleção dos estudos

Quadro 1 – Critérios de pontuação para avaliar a qualidade metodológica

Propriedade psicométrica	Critério de pontuação
Validade	Validade de conteúdo*



	Nível 1: Muito bom (++): O método para questionar os participantes e profissionais sobre a relevância e compreensibilidade foi amplamente utilizado e bem justificado; Todas as atividades físicas relevantes foram incluídas; A duração, frequência e intensidade foram abordadas e apropriadas para o construto a ser medido, assim como o tempo de recordação; As opções de resposta usadas foram apropriadas; O índice de validade de conteúdo (IVC) ou o percentual de concordância entre juízes foi avaliado.
	Nível 2: Adequado (+): O estudo assumiu um método para questionar os participantes e profissionais sobre a relevância e compreensibilidade que foi apropriado, mas não claramente descrito; A duração, frequência e intensidade foram abordadas e apropriadas para o construto a ser medido assim como o tempo de recordação; As opções de respostas usadas foram apropriadas.
	Nível 3: Duvidoso (-): Não ficou claro se os pacientes e profissionais foram questionados sobre a relevância e compreensibilidade dos conteúdos; A duração, frequência e intensidade não foram abordadas e apropriadas para o construto a ser medido assim como o tempo de recordação; As opções de respostas usadas não foram apropriadas
	Nível 4: Inadequado: (--): Não houve qualquer citação sobre a relevância e compreensibilidade.
	Validade de critério#
	Validade transcultural**
	Nível 1: Muito bom (++): Foram fornecidas provas de que a amostra era semelhante à população do questionário original; Houve método de tradução; A tradução foi realizada por especialistas, sendo mais de um tradutor independente; A tradução foi revista pelo comitê de especialistas; A amostra do pré-teste foi descrita adequadamente.
	Nível 2: Adequado (+): Foi declarado mas não comprovado que a amostra era semelhante à população do estudo original; Foi assumido que abordagem estatística foi apropriada, mas não houve descrição explícita.
	Nível 3: Duvidoso (-): Não ficou claro se as amostras eram semelhantes; Não foi definida qual abordagem foi utilizada e se era apropriada, como também os métodos.
	Nível 4: Inadequado (--): Nada relatado
Confiabilidade#	Nível 1: Muito bom (++): Houve intervalo de tempo adequado entre o teste e o reteste (2 dias a 2 semanas) e relato de ICC, kappa ou concordância.
	Nível 2: Adequado (+): O intervalo de tempo foi inadequado entre teste e reteste (> 2 semanas) e relato de ICC, kappa ou concordância; ou um intervalo de tempo adequado entre o teste e o reteste (2 dias a 2 semanas) e relato da correlação de Pearson/Spearman.
	Nível 3: Duvidoso (-): O intervalo de tempo foi inadequado entre teste e reteste (> 2 semanas) e relato de correlação de Pearson/Spearman.

* Terwee et al.²³, # Sattler et al.²¹; ** Mokkink et al.²⁴. # ver critérios da validade de critério no suplemento 2;

Tabela 1 — Parâmetros de validade de conteúdo, transcultural e de critério de questionários de atividade física durante a gestação

Validade de conteúdo e transcultural					Validade de critério			
Versões do PPAQ [#]	Validade de conteúdo	Qualidade dos estudos de validade de conteúdo	Validade transcultural	Principais Resultados	População do estudo (n =)	Medida de comparação; Tipo; Pontos de cortes ^a ; Local de uso; Período; Época; Dia válido	Resultados por intensidade de AF	Qualidade do estudo*
Versão turca ¹³	Comitê de especialistas piloto: (n [#] =10)	+	++	O item “cortar grama” foi substituído por “jardinagem”.	204	IPAQ ^{&} Saglam et al. ⁶²	Total: r = 0,67 Leve: r = 0,35 Intensa: r = 0,42 Sedentária: r = 0,65	3- 3- 3- 3-
					85	Pedômetro Cintura 7 dias		3-
						Acelerômetro (uniaxial) Hendelman et al. ⁶³	Total: rs = 0,27 Leve: rs = 0,27	2- 2-
Versão inglesa ²⁰	Recordatórios de 24 horas	+			54	Swartz et al. ⁶⁴ Freedson et al. ⁶⁵ Quadril direito 7 dias Época=1 min 8 horas/dia	Moderada: p = 0,38 Intensa: rs = 0,37 Sedentária: rs = -0,10 Doméstica: rs = -0,04 Ocupacional: rs = 0,16 Esporte: rs = 0,48 Sedentária: rs = -0,10	2- 2- 2-
Versão vietnamita ²⁶	Comitê de especialistas piloto: (n [#] =5)	+	++	O item “cortar grama” foi excluído e o item “andar de carro” foi substituído por “andar de bicicleta ou moto”.	59	Pedômetro Schneider et al. ⁶⁶ Cintura 14 dias 10 dias válidos	rs = 0,29	3-
Versão japonesa ²⁸	Comitê de especialistas piloto: (n [#] =10)	+	++	Foi adicionado o item “andar de bicicleta”	58	Swartz et al. ⁶⁴	1 ^a medida Total: rs = 0,02; rs = 0,35;	3-; 3-;
						Freedson et al. ⁶⁵	Acelerômetro (uniaxial) Leve ^a : rs = 0,01;	3-; 3-;
						Quadril direito 10 a 12 dias	Hendelman et al. ⁶³ Moderada ^a : rs = -0,09; rs = 0,38; rs = 0,30 Intensa ^a : rs = -0,25; rs = -0,21;	3-; 3-; 3-; 3-;
						Época=1 min 10 horas/dia rs = -0,23; Sedentária ^a : rs = 0,20; rs = -0,30; rs = 0,13;	3-; 3-; 3-;	
						2 ^a medida Total: rs = -0,05;	3-;	

						rs = 0,29;	3-;
						rs = 0,24	3-;
						Leve ^a :	
						rs = -0,01;	3-;
						rs = 0,13;	3-;
						rs = 0,13	3-;
						Moderada ^a :	
						rs = -0,11;	3-;
						rs = 0,38;	3-;
						rs = 0,23	3-;
						Intensa ^a :	
						rs = -0,25;	3-;
						rs = -0,21;	3-;
						rs = -0,23	3-;
						3 ^a medida	
						Total ^a :	
						rs = 0,04;	3-;
						rs = 0,13;	3-;
						rs = 0,20	3-;
						Leve ^a :	
						rs = 0,09;	3-;
						rs = 0,05;	3-;
						rs = 0,12	3-;
						Moderada ^a :	
						rs = -0,03;	3-;
						rs = 0,15	3-;
						rs = 0,12	3-;
						Intensa ^a :	
						rs = -0,18;	3-;
						rs = -0,09;	3-;
						rs = -0,18	3-;
						Sedentária ^a :	
						rs = 0,20;	3-;
						rs = -0,12;	3-;
						rs = 0,03;	3-;
Versão brasileira ²⁹	Comitê de especialistas piloto: (n [#] =20)	+	++	Foram excluídas a primeira e a segunda questão			
Versão chinesa ³⁰	Parecer de especialistas e estudos-piloto	+	++	-	125	Acelerômetro (uniaxial) Kumahara et al. ⁶⁷	Total: rs = 0,35 2-
						Quadril direito 7 dias 10 horas/dia	Leve: rs = 0,33 2- Intensa: rs = 0,15 2- Moderada: rs = 0,19 2-
Versão polonesa ³¹	Não específica	-	Sim	O item "cortar grama" foi excluído.			
Versão espanhola ³²	Comitê de especialista piloto: (n=16)	+	++	Sem alterações	208	Monitor multi-sensor Sensewear Mini Armband (SWA; Body media Inc, Pittsburgh, USA) Braço esquerdo 8 dias	AF total: rs = 0,004* 3- AFMV [%] : rs = 0,82 3+ Sedentária: rs = 0,001 3- Leve: rs = 0,07/ 0,122 3- Moderada: rs = 0,78/-0,019 3+ Intensa: rs = 0,24 3-
Versão dinamarquesa ³³	Comitê de especialista piloto: (n = 10)	+	++	Dois itens adicionais para ciclismo foram incluídos.			

				Três itens sobre passar tempo no computador foram mesclados em um único item.					
				O item "Cuidar de um idoso" foi considerado irrelevante num contexto dinamarquês e foi removido.					
Versão árabe ³⁴	Painel de especialista piloto: (n ^o =30) gestantes	+	++	Todos os itens foram considerados relevantes					
								Total: rs = 0,54	3-
								Sedentária rs = 0,86**	3+
								Leve = 0,27**	3-
								Moderada: rs = 0,49**	3-
								Intensa: rs = 0,27**	3-
								Doméstica: rs = 0,35**	3-
								Ocupacional; rs = 0,50**	3-
								Locomoção: rs = 0,77**	3+
								Lazer: rs = 0,73**	3+
Versão sérvia ³⁵	Painel de especialista piloto: (n ^o =15) gestantes	+	++	A medida inglesa (galões) foi substituída por "litros". O item "cortar grama" e as duas questões abertas foram excluídos;	150	IPAQ ^{&}			
Versão árabe ³⁶	Comitê de especialista piloto: (n ^o =45) gestantes	+	++	As três primeiras questões e "cortar grama" foram excluídos					
								Acelerômetro (triaxial)	Total: r = - 0,030
								Pontos de corte: Freedson et al. ⁶⁵	Leve: r = 0,149
								7 dias	Moderada: r = 0,003 (p = 0,961)
								Época= 1 min / 480 min /dia	Intensa: r = 0,106 (p = 0,11)
									Sedentária: r = 0,081 (p = 0,22)
Versão portuguesa ³⁷	Comitê de especialista piloto: (n ^o =6) gestantes	+	++	Foram feitas pequenas alterações semânticas	226				
Versão turca ³⁸	Painel de especialista piloto (n= 20)	+	++	Teve pequenas alterações nas medidas métricas.	328	IPAQ ^{&}			
									AF total:
									r ₁ = 0,17;
									r ₂ = 0,17;
									r ₃ = 0,21;
									Leve:
									r ₁ = 0,32;
									r ₂ = 0,26;
									r ₃ = 0,29;
Versão espanhola ³⁹	Comitê de especialista piloto (n ^o =58) gestantes	+	++	Substituição de alguns itens mais apropriados para gestantes espanholas					
Versão dinamarquesa ⁴⁰	Painel de especialista piloto:	+	++	Duas questões sobre o uso da TV e do iPad					

	(n [#] =10) gestantes			foram combinadas em uma questão. Uma questão de cortar grama foi excluída.				
Versão coreana ⁴¹	Painel de especialista CVR piloto= (n =3)	+	++	Não houve alterações.	363	GPAQ###	AF total: r = 1 Sedentária: r = -0,92* Leve: r = 0,84* Moderada: r = 0,82* Intensa: r = 0,06	3+ 3+ 3+ 3-
Versão grega ⁴³	Painel de especialista piloto = (n=46) IVC	+	Sim	Não houve alterações				
Versão persa ⁴²	Piloto= (n=20) gestantes. Comitê de especialista. RVC ^{&} IVC [*]	+	++	Pequenas alterações em alguns itens; as questões sobre jardinagem foram combinadas em um único item				

PPAQ = Questionário de Atividade Física na Gravidez; ⁵ IPAQ = Questionário Internacional de Atividade física; ### Questionário Global de Atividade física; n[#] = número da amostra, em relação aos critérios descritos; [&] RVC = Razão da validade de conteúdo; ⁹ qualidade metodológica dos estudos de acordo com o quadro 1; ^{*} IVC = índice de validade de conteúdo; ^a Correlação de Spearman usando os pontos de corte: Hendelman et al.⁶¹, Swartz et al.⁶² e Freedson et al.⁶³; Saglam et al.⁶⁰; rs = coeficiente de correlação de Spearman; r = coeficiente de correlação de Pearson; AF = atividade física; p = p valor; * significância estatística: p < 0,05; ** significância estatística: p < 0,01; AFMV% = atividade física moderada a vigorosa.

Tabela 2 – Parâmetros de confiabilidade de Questionários de Atividade física durante a gestação.

Versões do PPAQ [#]	População do estudo (n=)	Recall (dias)	Confiabilidade	Qualidade de evidência
Versão turca ¹⁴	204	7 dias	Total: ICC = 0,95	1+
			Leve: ICC = 0,93	1+
			Moderada: ICC = 0,96	1+
			Vigorosa: ICC = 0,98	1+
			Domicílio: ICC = 0,96	1+
			Ocupacional: ICC = 0,99	1+
			Esportes/exercício: ICC = 0,92	1+
Sedentária: ICC = 0,96	1+			
Versão polonesa ⁸	164	14 dias	Total: ICC = 0,77	1+
			Leve: ICC = 0,87	1+
			Moderada: ICC = 0,85	1+
			Vigorosa: ICC = 0,86	1+
			Domicílio: ICC = 0,89	1+
			Ocupacional: ICC = 0,86	1+
			Esportes: ICC = 0,83	1+
Locomoção: ICC = 0,59	1-			
Inatividade: ICC = 0,76	1+			
Versão inglesa ²⁰	54	7 dias	AF Total: ICC = 0,78	1+
			Inatividade: ICC = 0,79	1+
			leve: ICC = 0,78	1+
			moderada: ICC = 0,82	1+
			intensa: ICC = 0,81	1+
			Doméstica: ICC = 0,86	1+
			Ocupacional: ICC = 0,93	1+
Lazer/esporte: ICC = 0,83	1+			
Versão vietnamita ²⁶	60	14 dias	Total: ICC = 0,88	1+
			Leve: ICC = 0,88	1+
			Moderada: ICC = 0,90	1+
			Vigoroso: ICC = 0,87	1+
			Domicílio: ICC = 0,92	1+
			Ocupacional: ICC = 0,90	1+
			Esportes: ICC = 0,93	1+
Sedentária: ICC = 0,94	1+			
Esporte: ICC = 0,59	1+			
Versão chinesa ³⁰	125	7 dias	Total: ICC = 0,77	1+
			Leve: ICC = 0,75	1+
			Moderada: ICC = 0,59	1-
			Vigorosa: ICC = 0,28	1-
			Domicílio: ICC = 0,74	1+
			Ocupacional: ICC = 0,75	1+
			Esporte: ICC = 0,34	1-
Inatividade: ICC = 0,76	1+			
Versão polonesa ³¹	64	7 dias	Total: ICC = 0,75	1+
			Leve: ICC = 0,53	1-
			Moderada: ICC = 0,85	1+
			Vigorosa: ICC = 0,86	1+
			Domicílio: ICC = 0,89	1+
			Ocupacional: ICC = 0,86	1+
			Esportes: ICC = 0,83	1+
Locomoção: ICC = 0,59	1-			
Inatividade: ICC = 0,70	1+			
Versão espanhola ³²	109	8 dias	Total: ICC = 0,90	1+
			Leve: ICC = 0,88	1+
			Moderada: ICC = 0,96	1+
			Vigorosa: ICC = 0,92	1+
			Domicílio: ICC = 0,90	1+
			Transporte: ICC = 0,65	1-
			Ocupacional: ICC = 0,98	1+
Esportes: ICC = 0,94	1+			



			Total: ICC = 0,78	1+
			Leve: ICC = 0,78	1+
			Moderada: ICC = 0,80	1+
			Vigorosa: ICC = 0,61	1-
			Domicílio: ICC = 0,85	1+
			Transporte: ICC = 0,59	1-
			Ocupacional: ICC = 0,15	1-
			Esportes: ICC = 0,70	1+
			Inatividade: ICC = 0,65	1-
			Total: ICC = 0,93	1+
			Domicílio ICC = 0,76	1+
			Ocupacional: ICC = 0,81	1+
			Esportes: ICC = 0,84	1+
			Locomoção: ICC = 0,77	1+
			Inatividade: ICC = 0,78	1+
			Total: ICC = 0,77	1+
			Leve: ICC = 0,76	1+
			Moderada: ICC = 0,76	1+
			Vigorosa: ICC = 0,70	1+
			IPAQ	
			AF total: ICC = 0,88	1+
			Leve: ICC = 0,96	1+
			Moderada: ICC = 0,91	1+
			Intensa: ICC = 0,74	1+
			Sedentária: ICC = 0,90	1+
			Doméstica: ICC = 0,92	1+
			Ocupacional: ICC = 0,88	1+
			Total: ICC = 0,79	1+
			Leve: ICC = 0,71	1+
			Moderada: ICC = 0,84	1+
			Vigorosa: ICC = 0,75	1+
			Domicílio: ICC = 0,86	1+
			Transporte: ICC = 0,75	1+
			Ocupacional: ICC = 0,73	1+
			Esportes: ICC = 0,78	1+
			Inatividade: ICC = 0,61	1+
			Guttman® = 0,84.	3-
			Total: ICC _p = 0.55; ICC _s = 0.81 ICC _t = 0.83	1-; +1; +1
			Sedentária: ICC _p = 0.63; ICC _s = 0.82; ICC _t = 0.76	1-; +1; +1
			Leve: ICC _p = 0.75; ICC _s = 0.77 ICC _t = 0.87	1+; 1+; 1+
			AFMV®: ICC _p = 0.70; ICC _s = 0.81 ICC _t = 0.92	1+; 1+; 1
			Domicílio ICC = 0,72	1+
			Ocupacional ICC = 0,88	1+
			doméstica ICC = 0,88	1+
			esportes ICC = 0,86	1+
			lazer ICC = 0,93	1+

PPAQ = Pregnancy Physical activity Questionnaire; ICC=coeficiente de correlação intraclasse; ICC₁=semana para intervalo de uma semana; ICC₂=semanas para intervalo de duas semanas; qualidade metodológica conforme o quadro 1, @= Coeficiente de confiabilidade half-split de Guttman. ICC_p= coeficiente de correlação intraclasse; 1° trimestre; ICC_s= coeficiente de correlação intraclasse; no 2° trimestre; ICC_t= coeficiente de correlação intraclasse no 3° trimestre; &= AFMV= atividade física moderada a vigorosa

Tabela 3 – Perfil de evidência GRADE: Propriedades de medidas dos questionários de AF para a avaliação de AF total, leve e moderada

Confiabilidade								
Questionário	Intensidade da AF	Resultados	Número de estudo (n*)	Risco de Bia	Imprecisão	Indiretividade	Inconsistência	Qualidade de evidência
PPAQ_ Versão Inglesa ²⁰	Total	+	2(54)	Sério	Nenhum	Nenhum	Nenhum	Moderada
	Leve	+	2(54)	Sério	Nenhum	Nenhum	Nenhum	Moderada
	Moderada	+	2(54)	Sério	Nenhum	Nenhum	Nenhum	Moderada
PPAQ_ Versão Inglesa ¹⁴	Total	+	2 (50)	Nenhum	Nenhum	Nenhum	Nenhum	Alta
	Leve	+	2 (50)	Nenhum	Nenhum	Nenhum	Nenhum	Alta
	Moderada	+	2 (50)	Nenhum	Nenhum	Nenhum	Nenhum	Alta
PPAQ_ Versão Vietnamita ²⁶	Total	+	60	Sério	Nenhum	Nenhum	-	Moderada
	Leve	+	60	Sério	Nenhum	Nenhum	-	Moderada
	Moderada	+	60	Sério	Nenhum	Nenhum	-	Moderada
PPAQ_ Versão Japonesa ²⁷	Total	+	(58)	Sério	Nenhum	Nenhum	-	Moderada
	Leve	+	(58)	Sério	Nenhum	Nenhum	-	Moderada
	Moderada	+	(58)	Sério	Nenhum	Nenhum	-	Moderada
PPAQ_ Versão turca ¹³	Total	+	2(204)	Sério	Nenhum	Nenhum	Nenhum	Moderada
	Leve	+	2(204)	Sério	Nenhum	Nenhum	Nenhum	Moderada
	Moderada	+	2(204)	Sério	Nenhum	Nenhum	Nenhum	Moderada
PPAQ_ Versão turca ²⁸	Total	+	2 (90)	Sério	Nenhum	Nenhum	Nenhum	Moderada
	Leve	+	2 (90)	Sério	Nenhum	Nenhum	Nenhum	Moderada
	Moderada	+	2 (90)	Sério	Nenhum	Nenhum	Nenhum	Moderada
PPAQ_ Versão Chinesa ³⁰	Total	+	1 (125)	Sério	Nenhum	Nenhum		Moderada
	Leve	+	1 (125)	Sério	Nenhum	Nenhum		Moderada
	Moderada	-	1 (125)	Sério	Nenhum	Nenhum		Moderada
PPAQ_ Versão Espanhola ³²	Total	+	1 (109)	sério	Nenhum	Nenhum	-	Moderada
	Leve	+	1 (109)	sério	Nenhum	Nenhum	-	Moderada
	Moderada	+	1 (109)	sério	Nenhum	Nenhum	-	Moderada
PPAQ_ Versão Sérvia ³⁵	Total	+	30	Muito Sério	Muito sério	Nenhum	-	Muito baixa
PPAQ_ Versão portuguesa ³⁷	Total	+	184	Nenhum	Nenhum	Nenhum	-	Alta
	Leve	+	184	Nenhum	Nenhum	Nenhum	-	Alta
	Moderada	+	184	Nenhum	Nenhum	Nenhum	-	Alta
PPAQ_ Versão coreana ⁴¹	Total	+	187	Muito sério	Sério	Nenhum	-	Muito Baixa
Validade								
PPAQ_ Versão Inglesa ²⁰	Total	-	2 (54)	Sério	Sério	Nenhum	Nenhum	Baixo
	Leve	-	2 (54)	Sério	Sério	Nenhum	Nenhum	Baixo
	Moderada	-	2 (28)	Muito sério	Muito sério	Nenhum	Nenhum	Muito baixo
PPAQ_ Versão Vietnamita ²⁶	AF Total	-	59	Muito sério	Sério	Nenhum		Muito baixo



PPAQ_Versão Japonesa ²⁷	Total	-	54	Muito sério	Sério	Nenhum	-	Muito baixo
	Leve	-	54	Muito sério	Sério	Nenhum	-	Muito baixo
	Moderada	-	54	Muito sério	Sério	Nenhum	-	Muito baixo
PPAQ_Versão turca ¹³	Total	+	2 (204) (85)	Muito sério	Nenhum	Nenhum	Nenhum	Baixo
	Leve	-	2 (204)	Muito sério	Nenhum	Nenhum	Nenhum	Baixo
	Moderada	-	2 (204)	Muito sério	Nenhum	Nenhum	Nenhum	Baixo
PPAQ_Versão turca ¹³	Total	-	2 (328)	Muito sério	Nenhum	Nenhum	Nenhum	Baixo
	Leve	-	2 (328)	Muito sério	Nenhum	Nenhum	Nenhum	Baixo
PPAQ_Versão Chinesa ³⁰	Total	-	125	Sério	Nenhum	Nenhum		Moderada
	Leve	-	125	Sério	Nenhum	Nenhum		Moderada
	Moderada	-	125	Sério	Nenhum	Nenhum		Moderada
PPAQ_Versão Espanhola ³²	Total	-	208	Muito sério	Nenhum	Nenhum		Baixo
	Leve	-	208	Muito sério	Nenhum	Nenhum		Baixo
	Moderada	+	208	Muito sério	Nenhum	Nenhum		Baixo
PPAQ_Versão portuguesa ³⁷	Total	-	226	Sério	Nenhum	Nenhum		Moderada
	Leve	-	226	Sério	Nenhum	Nenhum		Moderada
	Moderada	-	226	Sério	Nenhum	Nenhum		Moderada
PPAQ_Versão Sérvia ³⁵	Total	-	150	Muito sério	Nenhum	Nenhum	-	Baixo
	Leve	-	150	Muito sério	Nenhum	Nenhum	-	Baixo
PPAQ_Versão coreana ⁴¹	Total	+	363	Muito sério	Nenhum	Nenhum		Baixo
	Leve	+	363	Muito sério	Nenhum	Nenhum		Baixo
	Moderada	+	363	Muito sério	Nenhum	Nenhum		Baixo
PPAQ_Versão Inglesa ¹⁴	Total	-	2(50)	Sério	Sério	Nenhum	Nenhum	Baixo
	Leve	-	2(50)	Sério	Sério	Nenhum	Nenhum	Baixo
	Moderada	+	2(50)	Sério	Sério	Nenhum	Nenhum	Baixo

Os resultados são demonstrados como propriedade psicométrica suficiente (+) ou insuficiente (-) de acordo com os critérios de pontuações das tabelas.