# Dez anos de Backpei: uma análise a partir da caracterização, representação e procedimentos

# INTRODUÇÃO

A medição é uma forma importante de produção de conhecimento dentro da ciência da saúde o que confere uma importância ao processo de desenvolvimento de sistemas de medição dentro do campo. Uma parte essencial para que estes sistemas sejam aplicados nas pesquisas é a verificação da sua adequação a partir de critérios científicos. Porém, ainda não está bem estabelecido dentro da ciência da saúde quais aspectos são determinantes na adequação de um sistema de medição e como avalia-los (Beraldo; Silva; Candotti, 2022; Lorente *et al.*, 2020; Mokkink, Lidwine B. *et al.*, 2010; Terwee *et al.*, 2016).

Ainda assim, várias pesquisas se dedicam ao desenvolvimento e avaliação de sistemas de medição que são posteriormente aplicados nas pesquisas dentro da ciência da saúde contribuindo para a produção de conhecimento do campo. A esse conjunto de pesquisas estamos chamando de clinimetria, que, embora não seja uma disciplina formalizada no campo da ciência da saúde, apresenta um significativo volume de estudos e grande importância. Assim, o estudo dos métodos e procedimentos adotados na clinimetria, e os resultados obtidos na aplicação dos sistemas de medição desenvolvidos e avaliados, é uma maneira de se conhecer os métodos para determinar a adequação dos sistemas de medição.

Uma forma de realizar este estudo da Clinimetria é a partir do uso da filosofia e da história como um complemento à ciência normal, a chamada ciência complementar (Chang, 2004). Neste artigo propomos utilizar a estrutura epistemológica para os sistemas de medição proposta por Nancy Cartwright e colaboradoras que, como apontada pelas autoras, pode ser aplicada à ciência da saúde e, consequentemente, à clinimetria (Bradburn; Cartwright; Fuller, 2016; Cartwright; Runhardt, 2018). Segundo às autoras os sistemas de medição em saúde exigem que:

1. Definamos o conceito ou quantidade, identificando seus limites e determinando quais características pertencem a ele e quais não (caracterização); 2. Definamos um sistema métrico que represente apropriadamente o conceito ou quantidade (representação); 3. Formulemos regras para aplicar o sistema métrico a alvos para produzir os resultados da medição (procedimentos) (Bradburn; Cartwright; Fuller, 2016, p. 3).

Como alvo de análise propomos analisar pela ciência complementar o desenvolvimento e avaliação do *Back Pain and Body Posture Assessment Intrument* (BackPEI). Este sistema de medição foi desenvolvido em 2013 e é adequado à proposta deste artigo por se tratar de um instrumento que passou por estudos de desenvolvimento (Candotti *et al.*, 2018; Da Rosa *et al.*, 2022; Noll *et al.*, 2013a), alterações (Candotti *et al.*, 2018, 2023) e avaliação (Antoniolli *et al.*, 2015; Gökşen; Kocaman; Yıldırım, 2023), possui diferentes versões desenvolvidas por grupos de pesquisa distintos (Gençbaş; Bebiş, 2019; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a) e foi aplicado em diversas pesquisas da ciência da saúde (Aimi *et al.*, 2019; Akbari-Chehrehbargh; Tavafian; Montazeri, 2020; Da Rosa *et al.*, 2018; Galmes-Panades; Borràs; Vidal-Conti, 2023; Melo-Marins; Carvalho; Gomes, 2015; Niaradi; Niaradi; Gasparetto, 2024; Noll *et al.*, 2013b, 2016, 2017; Ozdemir *et al.*, 2021; Vidal-Conti *et al.*, 2023; Vitman *et al.*, 2022).

Portanto, a presente revisão de escopo pretende analisar o processo de desenvolvimento e avaliação do BackPEI a partir da estrutura epistemológica para sistemas de medição na saúde: caracterização, representação e procedimentos.

# METODOLOGIA

Foi utilizada a metodologia da revisão de escopo (Peters *et al.*, 2020) e o guia para reportar estudos PRISMA-ScR (Tricco *et al.*, 2018). Foi mantido o registro de toda a produção e alteração de documentos na condução da pesquisa a partir de um gestor de versionamento de arquivos, os registros estão públicos e podem ser consultados em <https://github.com/lucasmberaldo/10-anos-de-back-pei>.

Foram estabelecidos como critério de elegibilidade: publicações a partir de 2013 (ano da publicação da primeira versão do BackPEI); artigos originais publicados em periódicos da ciência da saúde (excluindo-se revisões, por exemplo); a utilização de alguma versão do BackPEI dentro da metodologia do estudo; e estudos de clinimetria (desenvolvimento e avaliação de sistemas de medição em saúde).

Inicialmente partiu-se de um conjunto de 32 fontes iniciais conhecidas por nós que cumpriam quase todos os critérios de elegibilidade, excetuando apenas o referente à clinimetria por compreenderem também estudos de aplicação do BackPEI. Estas fontes iniciais foram utilizadas para a confecção das chaves e controle da abrangência da busca. Foram utilizadas 3 bases para a sequência da busca das fontes: Pubmed, Scopus e Lilacs a partir da chave: "Sırt Ağrısı ve Vücut Duruşu Değerlendirme Aracı’nın" OR "BackPEI" OR "BackPEI-A" OR "BackPEI-CA" OR "Back Pain and Body Posture Evaluation Instrument". Também foram identificadas entre as 32 fontes iniciais quais eram estudos de clinimetria do BackPEI, ou seja, de desenvolvimento ou avaliação do sistema de medição. Realizamos então uma ampla busca dos estudos que realizavam citações a estas pesquisas clinimétricas utilizando o Web of Science, Scopus, Pubmed, Semantic Scholar e Research Gate. A pesquisa de fontes a partir das chaves de busca e das citações dos artigos de clinimetria do BackPEI aconteceram entre 20/09/2024 e 30/09/2024.

As fontes iniciais e os artigos encontrados a partir das buscas foram importados para a plataforma Rayyan sendo feita uma primeira remoção das duplicatas. A seleção dos artigos a partir dos critérios de elegibilidade foi realizada de forma cegada por dois avaliadores independentes. Ao final de cada etapa o cegamento era quebrado e os casos de dissenso eram resolvidos em reunião entre os avaliadores e, em caso de persistência do dissenso, um terceiro avaliador determinaria a inclusão ou não de uma fonte. Um dos avaliadores conferia a concordância a cada sessão de seleção, sendo estabelecido que caso esta fosse inferior a 75% o processo seria interrompido. Neste momento, conforme indicado para revisões de escopo (Peters *et al.*, 2020), o cegamento seria quebrado e a equipe se reuniria para discutir as discrepâncias e, caso necessário, reformular os critérios de elegibilidade.

A primeira seleção foi realizada a partir da leitura dos títulos e resumos. Nesta fase foram mantidos os artigos de revisão e aqueles que não se tratavam pesquisas de clinimetria. A segunda seleção aconteceu a partir do acesso aos artigos na íntegra. Foi investigado entre os artigos de revisão se alguma das fontes utilizadas neles poderiam servir como fontes desta revisão de escopo e estas seriam incluídas no processo de seleção, posteriormente as revisões foram removidas. Nesta fase, novamente, os estudos que não eram de clinimetria foram mantidos. Na última etapa, os estudos restantes foram divididos entre aqueles que são de aplicação do BackPEI o os que são de avaliação ou desenvolvimento de alguma versão do BackPEI (clinimetria).

Essas fontes foram importadas para o *software* Zotero, para a gestão de referências, onde já se encontravam as 32 fontes iniciais. Foi identificado se essas fontes iniciais se encontravam entre as fontes selecionadas e, mais uma vez, foram verificadas possíveis duplicatas. A classificação entre estudos de aplicação e clinimetria foram verificadas por um avaliador, sendo que aqueles referentes a clinimetria foram selecionados para compor o conjunto de fontes desta revisão.

Para a extração dos dados das fontes foi confeccionada uma planilha utilizando o Microsoft Excel sendo identificados os autores, o ano, o objetivo do estudo, cada procedimento utilizado no estudo, o método aplicado a cada procedimento, o resultado da aplicação de cada método e a conclusão do estudo.

Além disso, a partir da estrutura epistemológica proposta por Cartwright e coloboradoras (Bradburn; Cartwright; Fuller, 2016; Cartwright; Montuschi, 2018) foi realizada uma análise de conteúdo em três etapas. Inicialmente foram identificados e classificados trechos das fontes a partir de quatro categorias, os três aspectos da medição presentes na estrutura epistemológica (caracterização, representação e procedimentos) e o propósito do estudo. O último não é um aspecto presente na estrutura, mas as autoras apontam a sua importância, dado que os aspectos devem ser analisados a partir do propósito. Esses trechos foram registrados na mesma planilha. Na segunda etapa foi realizada uma síntese dos trechos selecionados na primeira etapa identificando como a caracterização, representação e procedimentos eram realizados em cada fonte, além do propósito do sistema de medição. Na terceira etapa, a partir das sínteses elaboradas, buscou-se a reposta para as seguintes questões: O que o BackPEI mede? Como são definidos os conceitos medidos pelo BackPEI? Como a caracterização é avaliada? Quais mudanças aconteceram no conceito? Qual é o resultado da medição pelo BackPEI? A medição pelo BackPEI resulta em quais tipos de dados? Como a representação é avaliada? Quais mudanças aconteceram na representação? Quais as operações do BackPEI? Como são avaliados os procedimentos? Quais mudanças aconteceram nos procedimentos?

# RESULTADOS

O processo de seleção das fontes de evidência está apresentado no diagrama da Figura 1. Inicialmente foram reunidas 381 possíveis fontes, após a remoção das duplicatas passou-se a seleção a partir dos títulos e resumos. Após a avaliação das primeiras 50 fontes por ambos os avaliadores foi verificada uma concordância de 62%, a seleção foi interrompida e realizada uma reunião entre a equipe onde as discrepâncias foram debatidas. Os critérios de elegibilidade foram mantidos e o processo de decisão foi refinado e alinhado. Após a reunião a concordância na seleção a partir da leitura dos títulos e resumos foi de 82% entre os avaliadores e na seleção com acesso aos artigos na íntegra foi de 100%. Assim, não foi necessária nova interrupção nesse processo devido à alta concordância entre os avaliadores.

Figura 1-Fluxograma do processo de seleção dos artigos

Diagrama

O conteúdo gerado por IA pode estar incorreto.

As razões para exclusão das possíveis fontes estão apontadas na Figura 1 sendo que algumas delas descumpriam mais de um critério de elegibilidade. Nós discriminamos os estudos excluídos que não utilizavam o BackPEI em três categorias: aqueles que avaliam desfechos distintos dos avaliados pelo BackPEI (desfechos); aqueles que utilizam instrumentos diversos do BackPEI (instrumentos); e aqueles que utilizam animais não-humanos como amostra (animais). Também apontamos qual o tipo de estudo, quando não se tratava de artigos originais. Destacamos que na leitura dos títulos e resumos dois protocolos de revisão foram excluídos por não se tratarem de estudos originais ou revisões. Não obstante, a equipe identificou que esses protocolos resultaram em dois artigos de revisão publicados e que ambos estavam incluídos entre as possíveis fontes, tendo sido avaliados.

Ao acessar os artigos na íntegra encontramos textos em português, inglês e turco. O último idioma não era compreendido por nenhum dos avaliadores, entretanto, optou-se por não o excluir. A equipe utilizou das ferramentas copilot e google tradutor para gerar uma versão em inglês e a utilizou para fazer a avaliação do artigo. Como ele foi incluído entre as fontes ao final da seleção utilizou-se da versão gerada pela equipe para as análises. Durante a fase de seleção a partir dos artigos na íntegra verificamos que dois artigos constantes nas revisões eram possíveis fontes, elas foram inseridas na plataforma Rayyan. Não obstante, nessa plataforma foi identificado que ambas as fontes já haviam sido avaliadas pela equipe se tratando, portanto, de duplicatas. Assim as 61 possíveis fontes foram exportadas do Rayyan e importadas para o Zotero onde verificou-se que todas as 32 fontes iniciais estavam presentes. Por fim restaram dez artigos após a identificação dos estudos de clinimetria e estes compõem as fontes dessa revisão (Tabela 1).

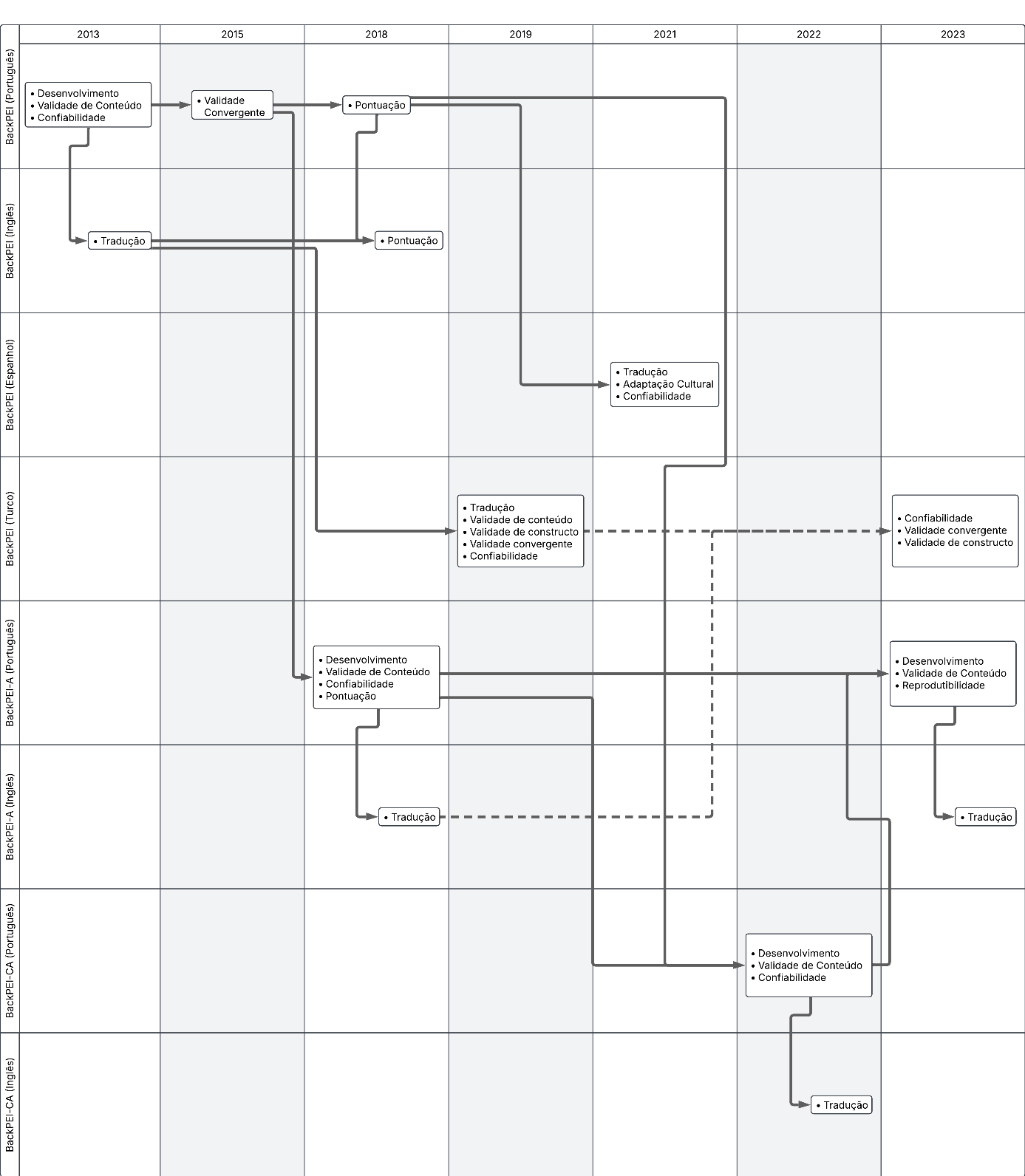
|  |  |  |  |
| --- | --- | --- | --- |
| **Fonte** | **Etapa** | **Método** | **Resultado** |
| Noll et al. (2013) | Decisões prévias baseados em artigos conhecidos previamente pelos autores | Inespecífico | Elaboração dos itens relativos à dor. Decisão de utilizar fotografias nos itens relativos aos hábitos posturais e ter questionários diferentes com base no gênero. |
| Pesquisa de campo para registrar os hábitos postuais mais comuns e seu modo de execução | Inespecífico | Posições identificadas como mais comuns foram consideradas para as fotografias |
| Revisão de literatura para identificar fatores de risco | Inespecífico | Elaboração dos itens relativos aos fatores de risco. |
| Validade de conteúdo | Painel com 8 especialistas avaliou: adequação ao avaliar AVDs a partir de fotografias; adequação ao avaliar dor nas costas e fatores associados; a clareza, facilidade de compreensão e aplicabilidade. Avaliação através de tabelas de frequência. | Recomendações: melhor estruturação das questões; melhorar a qualidade das imagens; alterar algumas imagens; preparar uma versão para cada gênero com posturas diferentes. Aprovação do sistema de medição na segunda rodada de consulta ao painel. |
| Avaliação qualitativa da compreensão do questionário | O questionário foi aplicado a 20 crianças em idade escolar e ocorreu uma avaliação qualitativa | Não havia necessidade de mudanças |
| Confiabilidade teste-reteste (reprodutibilidade através de um procedimento de teste e reteste) | O questionário foi aplicado duas vezes entre 260 indivíduos com intervalo de 7 dias. A avaliação ocorreu através do κ não-ponderado (≥0,5), CCI e teste dos postos de sinais de Wilcoxon (p>0,05). | Todos os itens apresentaram κ>0,5. A intensidade da dor obteve CCI=0,937 com diferença não-significativa segundo o teste dos postos de sinais de Wilcoxon (p=0,251). . |
| Tradução para o inglês | Tradução independente por 3 tradutores. Análise da equivalência semântica, cultural e conceitual por um painel de 2 especialistas. | Versão em inglês do questionário |
| Antoniolli et al. (2015) | Concordância e associação entre os resultados do BackPEI e LADy | Os hábitos posturais de 58 estudantes foram medidos através do LADy e do BackPEI a concordância foi avaliada através do κ, %C (>80%) e χ² (p<0,05) | Hábitos posturais: carregar mochila (%C 96,5%; κ=0,905 ;χ²=47,15 , p=0,001), pegar objeto do chão (%C 98,3%; κ=0,962 ;χ²=53,78 , p=0,001), sentar para escrever (%C 93,1%; κ=0,737 ;χ²=31,51 , p=0,001) e usar o computador (%C 84,5%; κ=0,483 ;χ²=14,757 , p=0,01). |
| Pivotto et al. (2018) | Pontuação do BackPEI | Inespecífico | Envolve apenas as questões referentes aos fatores de risco. Obtida pela soma dos pontos (máximo de 16 pontos). Quanto maior o valor, menor a exposição a fatores de risco. |
| Candotti et al. (2018) | Revisão de literatura e experiência das autoras para identificar fatores de risco para dor nas costas e pescoço entre adultos | Inespecífico | Alteração das questões originais e elaboração da versão inicial do questionário para adultos |
| Validade de conteúdo | Painel com 3 especialistas avaliou: adequação ao avaliar AVDs a partir de fotografias; adequação ao avaliar dor nas costas e no pescoço e fatores associados; a clareza, facilidade de compreensão e aplicabilidade. | Especialistas recomendaram pequenas alterações que foram realizadas. Na segunda rodada de consulta constatou-se que não seria preciso nenhuma alteração |
| Confiabilidade teste-reteste (Reprodutibilidade) | O questionário foi aplicado duas vezes entre 154 indivíduos com intervalo de 7 dias. A avaliação ocorreu através do κ (≥0,5), CCI2,2. | Todos os itens apresentaram κ>0,5. A intensidade da dor apresentou CCI=0,908 para dor nas costas e CCI=0,865, para dor no pescoço. |
| Tradução para o inglês | Tradução independente por 3 tradutores. Análise da equivalência semântica, cultural e conceitual por um painel de 2 especialistas. | Versão em inglês do questionário |
| Pontuação do BackPEI-A | Inespecífico | Envolve apenas as questões referentes aos fatores de risco. Obtida pela soma dos pontos (máximo de 10 pontos). Quanto maior o valor, menor a exposição a fatores de risco. |
| Gençbas e Bebis (2019) | Tradução para o turco | 1 tradutor realizou a tradução do inglês para o turco, outro tradutor avaliou a compreensão e gramática. | Versão preliminar do questionário em turco |
| Validade de Conteúdo | Painel com 5 especialistas verificaram a clareza, relevância e adequação de cada questão. Avaliação através do IVC. | O item referente a ler e/ou estudar na cama foi separada em 2: ler na cama; e estudar na cama. Obteve-se um IVC=0,908. Elaboração da versão final do questionário em turco do BackPEI. |
| Validade de linguagem | O questinário foi retraduzido para o inglês para verificar se houve alteração do significado | Não houve alteração do significado das questões |
|  | Avaliação da compreensão do questionário | O questionário foi aplicado entre 10 estudantes de perfil similar à população do estudo | Decidiu-se que é compreensível |
| Validade de constructo | O questionário foi aplicado entre 139 estudantes e foi avaliado através dos valores de KMO e do teste de esfericidade de Bartlett. | KMO=0,374, Bartlett (χ²=464,7, p=0,000), |
| Validade convergente (validade de critério) | O questionário foi aplicado entre 139 estudantes. Foi analisada a associação entre a intensidade da dor nas costas e a presença de dor nas costas através do ρ | ρ=-0,684 |
| Confiabilidade teste-reteste (Reprodutibilidade) | O questionário foi aplicado entre 139 estudantes com 2 semanas de intervalo. A avaliação ocorreu através do κ e testes dos postos de sinal de Wilcoxon. | κ entre 0,500 e 0,880. Sem diferença para a intensidade da dor (Wilcoxon) |
| Akbari-Chehrehbargh, Tavafian e Montazeri (2020) | Revisão de literatura | Inespecífico | Versão inicial do BABAQ com 55 questões para 5 construtos pré-definidos |
| Validade de conteúdo | Painel com 13 especialistas avaliou a utilidade e relevância de cada item. A avaliação aconteceu através do IVC (>0,7) e RVC (≥0,54). | 4 itens foram excluídos. Elaboração da versão intermediária do BABAQ. |
| Avaliação qualitativa da compreensão do questionário (Validade de face) | O questionário foi aplicado entre 6 alunas da 5ª série que avaliaram se conseguiam ler e entender as questões. | 2 itens foram excluídos. Elaboração da versão final do BABAQ. |
| Validade estrutural | O questionário do BABAQ foi aplicado entre 610 indivíduos. Foi realizada uma Análise Fatorial Confirmatória. A avaliação ocorreu através da χ²/gl (<5), CFI, RMSEA (<0,10), SRMR (<0,08) e correlação item-total | χ²/gl=3,51; ICA=0,97; RMSEA=0,091; SRMR=0,078; correlação item-total satisfatória; |
| Validade convergente | Variância Média Extraída (>0,5) | Habilidades=0,54; Conhecimento=0,73; autoeficácia=0,79; crenças=0,49 e comportamento=0,86 |
| Confiabilidade | O questionário foi aplicado duas vezes entre 50 indivíduos com intervalo de 2 semanas. A avaliação ocorreu através do α (>0,70), ICC (>0,70), EPM e MMD | α entre 0,93 e 0,97, ICC entre 0,76 e 0,83, EPM entre 0,72 e 7,08 e MMD entre 1,99 e 19,62 |
| Miñana-Signes et al. (2021) | Tradução | Um tradutor traduziu do português para o espanhol. Um segundo tradutor retraduziu do espanhol para o português. Um terceiro tradutor comparou a versão traduzida para o espanhol e retraduzida para o português. A versão preliminar foi estabelecida por consenso entre os três tradutores. | Nenhuma intervenção foi feita. Questionário retraduzido para o português era idêntico ao original em português. |
| Adaptação Cultural | Painel com 2 especialistas verificaram a equivalência da versão em espanhol quanto a semântica, idiomática, experiencial e conceitual. Erros de revisão foram corrigidos. Avaliação ocorreu através da concordância (≥80%). | A 2ª rodada já resultou em consenso, entretanto, acolheu-se novas sugestões feitas. Na 3ª rodada houve 100% de concordância em todos os itens. Não foram encontrados erros na revisão. |
| Confiabilidade | O questionário foi aplicado duas vezes entre 224 indivíduos com intervalo de 7 dias. A avaliação ocorreu através do ICC (≥0,75), teste dos postos de sinais de Wilcoxon e κ com %C (κ≥0,5 ou %C≥75%). | 2 itens com κ<0,5 e 4 itens com %C<75%, nenhum inferior em ambos. A intensidade da dor obteve ICC=0,928, sem diferença significativa segundo Wilcoxon. |
| Rosa et al. (2022) | Revisão de literatura para identificar como a dor nas costas é avaliada e fatores de risco | Inespecífico | Versão preliminar do questionário do BackPEI-CA |
| Validade de conteúdo | Painel com 8 especialistas analisaram: clareza, facilidade de compreensão e aplicabilidade geral das novas questões; se as novas questões permitiam a identificação do comportamento de uso de dispositivos móveis; se a nova versão avaliava adequadamente dor nas costas e no pescoço. A avaliação aconteceu através do IVC para cada questão (>0,8, preferencialmente >0,9). | Na 2ª rodada o ICV variou de 0,9 a 1,0. Elaboração da versão final do BackPEI-CA |
| Confiabilidade teste-reteste | O questionário foi aplicado duas vezes entre 105 indivíduos com intervalo de 7 dias no formato presencial e virtual. A avaliação ocorreu através do %C (>50%), κ (≥0,4) e CCI(2-way mixed effects absolute agrement) | Todos os itens apresentaram %C>50% e κ>0,4. A intensidade da dor apresentou CCI=0,828 para dor nas costas e CCI=0,824 para dor no pescoço. |
| Tradução para o inglês | Inespecífico | Versão em inglês do questionário do BackPEI-CA |
| Candotti et al. (2023) | Revisão de literatura para identificar as posturas adotadas ao utilizar dispositivos móveis | Inespecífico | Inclusão de questões novas |
| Validade de conteúdo | Painel com 8 especialistas avaliou: clareza, facilidade de compreensão e aplicabilidade das novas questões e novo design; se as novas questões permitem a identificação dos fatores comportamentais relativos ao uso de dispositivos móveis. A avaliação aconteceu através do IVC para cada questão (≥0,8). | Na 2ª rodada foram obtidos ICV=1 para novas questões e design e ICV = 0,87 para os fatores comportamentais. Elaboração da versão final do BackPEI-A |
| Avaliação da confiabilidade | O questionário foi aplicado duas vezes entre 194 indivíduos com intervalo de 7 dias no formato presencial (n=89) e virtual (n=105). A avaliação ocorreu através do %C (>50%), κ (>0,4) e CCI2,2, EPM, MMD | Todos os itens apresentaram %C>50% e κ>0,4. A intensidade da dor nas costas apresentou CCI=0,595, SEM=1,26cm e MMD=2,48cm. A intensidade de dor no pescoço apresentou CCI=0,479, SEM=1,68cm e MMD=3,10cm. |
| Tradução para o inglês | Inespecífico | Versão em inglês do questionário do BackPEI-A |
| Gökşen, Kocaman e Yildirim (2023) | Confiabilidade teste-reteste (Consistência interna - a parte do kappa e %C) | O questionário foi aplicado duas vezes entre 235 indivíduos com intervalo de 7 dias. A avaliação ocorreu através do κ (≥0,5), %C, EP, CCI (2-way random effects - absolute agreement), r, teste dos postos de sinais de Wilcoxon.0,5), %C e SE | Todos os itens apresentaram κ>0,5, %C e EP variaram respectivamente de 74,5% a 100% e 0,000 a 0,138. A intensidade da dor nas costas apresentou CCI=0,716. A intensidade de dor no pescoço apresentou CCI=0,944. A pontuação total do BackPEI apresentou CCI=0,800, r=0,809 e diferença significativa segundo o teste dos postos de sinais de Wilcoxon (p=0,009). |
| Validade convergente (validade concorrente) | O BackPEI-TR e o MHQ-TR foram aplicados entre 235 indivíduos. A avaliação ocorreu a partir do ρ. | ρ=0,725. |
| Validade de constructo | A partir da pontuação do BackPEI obtida entre 235 indivíduos foram determinados dois grupos: aqueles com valores acima do 73º percentil (G>73º); e aqueles com valores abaixo do 23º percentil (G<23º). Os valores dos grupos foram comparados através do teste t independente. | Diferença significativa segundo o teste t (p<0,001) |

Tabela 1 – Fontes incluídas na revisão

Entre as dez fontes, nove estudos são de desenvolvimento e/ou avaliação de alguma versão do BackPEI (Antoniolli *et al.*, 2015; Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Gökşen; Kocaman; Yıldırım, 2023; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a; Pivotto *et al.*, 2018), o estudo restante realiza o desenvolvimento e avaliação do BABAQ, um sistema de medição dedicado a avaliação de comportamentos de cuidados com a lombar (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020). Este estudo foi mantido, mesmo não tendo foco no BackPEI, pois trata-se de uma pesquisa de clinimetria e o BacKPEI foi utilizado no desenvolvimento do BABAQ.

Foram identificadas oito diferentes versões do BackPEI e seu desenvolvimento e avaliação estão expostos na Figura 2**.** Em 2013 foi realizado o desenvolvimento da versão original do BackPEI, em português, e realizada a tradução do questionário para o inglês (Noll *et al.*, 2013a). Essas versões ganham um sistema de pontuação em 2018 (Pivotto *et al.*, 2018). Originalmente o BackPEI se dedicava a dor lombar e fatores de risco e era destinado a crianças em idade escolar. Em 2018 também foi elaborado o BackPEI-A, destinado a adultos e que inclui, além da dor lombar e fatores de risco, a dor cervical e seus fatores risco, já contando com um sistema de pontuação. O BackPEI-A também foi desenvolvido em português e traduzido para o inglês (Candotti *et al.*, 2018). A partir da tradução em inglês do BackPEI original foi desenvolvida uma versão em turco em 2019 (Gençbaş; Bebiş, 2019) e a versão em espanhol é desenvolvida em 2021 a partir da versão em português do BackPEI já considerando o sistema de pontuação (Miñana-Signes *et al.*, 2021). Em 2022, tendo como base as versões em português do BackPEI e BackPEI-A foi desenvolvido um novo sistema de medição, o BackPEI-CA. Este é destinado a crianças e adolescentes e tem como objeto de medição as dores nas costas e pescoço e seus fatores de risco, incluindo o uso de dispositivos móveis e a aplicação remota do questionário. Esta nova versão foi desenvolvida em português e foi traduzida, tendo também uma versão em inglês (Da Rosa *et al.*, 2022). Posteriormente, em 2023, o BackPEI-A foi atualizado passando a abranger também o uso de dispositivos móveis como um fator de risco, além da possibilidade de aplicação digital (Candotti *et al.*, 2023).

Figura 2 - Linha do tempo do BackPEI.



Em 2023 é realizada uma avaliação do BackPEI entre pacientes com hérnia de disco (Gökşen; Kocaman; Yıldırım, 2023). Embora os autores indiquem que utilizaram a versão do BackPEI em turco no estudo são apontados elementos que não faziam parte dessa versão: a medição dos aspectos da dor no pescoço, e a pontuação de 0 a 10 para os fatores de risco. Portanto, presumimos que foi utilizada uma nova versão em turco traduzida a partir da versão em inglês do BackPEI-A desenvolvido em 2018. Entretanto o desenvolvimento dessa versão não é citado por Gökşen, Kocaman e Yıldırım (2023) e não a encontramos em nossa busca pelas fontes. Nós entramos em contato com os autores solicitando o questionário utilizado na pesquisa e fizemos uma busca não padronizada para encontra-lo, entretanto não obtivemos retorno dos autores e nem encontramos o questionário. Por não haver certeza sobre a origem do sistema utilizado nesta pesquisa optamos por representar a origem presumível e a alegada através de linhas tracejadas na Figura 2.

Para analisar os métodos de avaliação empregados realizamos uma padronização da nomenclatura, pois, alguns estudos indicavam a mesma metodologia a partir de nomes distintos. Para esse processo adotamos a nomenclatura a partir de dois critérios: maior frequência de uso e uso mais recente.

Os sistemas do BackPEI foram desenvolvidos principalmente a partir de revisões de literatura (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Noll *et al.*, 2013a), embora nenhum estudo apontou a metodologia para condução dessa revisão e elaboração dos procedimentos a partir dela.

A confiabilidade teste-reteste (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Gökşen; Kocaman; Yıldırım, 2023; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a) e validade de conteúdo (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Noll *et al.*, 2013a) foram as metodologias mais empregadas na avaliação do BackPEI. Estes método também foi empregado no desenvolvimento do BABAQ, sendo que Akbari-Chehrehbargh, Tavafian e Montazeri (2020), ao avaliar a confiabilidade, fizeram uma análise transversal através do alfa de Cronbach além da confiabilidade teste-reteste.

As fontes também apresentaram diferentes métodos e critérios de avaliação para a validade de conteúdo. Foi utilizado o Índice de Validade de Conteúdo com diferentes critérios: valores de 0,8 ou maiores (Candotti *et al.*, 2023); valores superiores a 0,8 (Da Rosa *et al.*, 2022); valores acima de 0,7 (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020); e mesmo sem algum critério (Gençbaş; Bebiş, 2019). A razão de validade de conteúdo também foi empregada, com ponto de corte de 0,54 (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020). Os estudos de Noll et al. (2013a) e Candotti et al. (2018) analisaram a validade de conteúdo, mas sem um critério ou uma medida.

Houve também alguma diferença entre as perguntas feitas ao painel que refletem as diferenças entre as diferentes versões do instrumento. Entretanto, a validade de conteúdo aplicada no BackPEI difere da aplicada ao BABAQ, pois parece estar mais voltada para aspectos operacionais do questionário como o uso de fotografias e a redação dos itens.

A confiabilidade teste-reteste apresentou uma variedade de diferença entre as fontes. A quantidade de indivíduos avaliado foi no mínimo 50 (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020) e no máximo 260 (Noll *et al.*, 2013a). Algumas fontes utilizaram um intervalo de 7 dias (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gökşen; Kocaman; Yıldırım, 2023; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a) entre as aplicações e outras 14 dias (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020; Gençbaş; Bebiş, 2019). A confiabilidade foi avaliada através de diversos testes e medidas estatísticas, com destaque para o kappa de Cohen (Candotti *et al.*, 2018, p. 201, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Gökşen; Kocaman; Yıldırım, 2023; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a) e o coeficiente de correlação intra-classe (CCI) (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020; Candotti *et al.*, 2018, p. 201, 2023; Da Rosa *et al.*, 2022; Gökşen; Kocaman; Yıldırım, 2023; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a). Embora 2 dos 6 estudos que empregaram o CCI não informaram qual o modelo utilizado (Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a).

## Caracterização

Todas as versões do BackPEI medem a dor em alguma região da coluna e seus fatores associados para alguma faixa etária. As versões originais (em português e inglês) (Noll *et al.*, 2013a), a versão em turco (Gençbaş; Bebiş, 2019) e a versão em espanhol (Miñana-Signes *et al.*, 2021) avaliam a dor nas costas e as demais versões, BackPEI-A (Candotti *et al.*, 2018, 2023) e BackPEI-CA (Da Rosa *et al.*, 2022), avaliam a dor nas costas e a dor no pescoço.

A região “das costas” corresponde a qualquer região torácica e lombar e a região “do pescoço” corresponde à coluna cervical. Os artigos citam que a dor é medida a partir de três elementos: a presença, a frequência e a intensidade da dor (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a). Entretanto, observando os itens do questionário, nota-se que também é medida a incapacidade devido à dor, como, por exemplo: no BackPEI original há o item “Esta dor nas costas impede ou impediu de realizar atividades como: brincar, estudar, praticar esportes...” (Noll *et al.*, 2013ª, Supplementary material 1, p. 3); no BackPEI-A, “Esta dor no pescoço impede ou impediu de realizar atividades como: trabalhar, ler, praticar esportes...” (Candotti *et al.*, 2023, p. 160) tendo também um pergunta semelhante se referindo à dor nas costas; e o BackPEI-CA tem duas perguntas relativas a incapacidade para cada região de dor (nas costas e no pescoço), “Você já teve que faltar aula por causa da dor no pescoço?” e “Você teve que parar de praticar esportes e/ou brincar por causa da dor nas costas?” (Da Rosa *et al.*, 2022, p. 4). Em todas as versões do BackPEI é estabelecido que a dor medida, em todos os seus elementos, se limita aos 3 meses anteriores.

Existem versões para o BackPEI destinadas à adultos (18 a 80 anos) (Candotti *et al.*, 2023) e a crianças e adolescentes em idade escolar (10 a 17 anos) (Da Rosa *et al.*, 2022; Miñana-Signes *et al.*, 2021). Nota-se que a faixa etária é determinante da caraterização pois os conceitos são diferentes para cada faixa etária. A principal marca disso é que os fatores de risco são diferentes entre adultos e as crianças e adolescentes.

Os fatores de riscos quem compõe todas as versões do BackPEI são inatividade física, frequência de atividade física, prática esportiva competitiva, tempo assistindo TV, tempo utilizando o computador, posição inadequada ao dormir, tempo de sono, postura inadequada ao escrever, postura inadequada ao sentar para conversar, postura inadequada ao utilizar o computador, postura inadequada para pegar um objeto do chão, presença de dor nos pais. Para os adultos, também é apontado o envolvimento em acidentes de carro como um fato de risco para a dor no pescoço. Todos estes fatores de risco são agrupados em um domínio denominado comportamental, que, às vezes, é dividido entre hábitos de vida e hábitos posturais. As versões mais recentes do BackPEI incluem o hábito de ler, estudar ou usar dispositivos móveis na cama, o tempo usando dispositivos móveis, a postura inadequada ao utilizar dispositivos móveis sentado e a postura inadequada ao utilizar dispositivos móveis em pé, tanto no BackPEI-A, destinado a adultos, quanto no BackPEI-CA, destinado a crianças e adolescentes em idade escolar. Exclusivamente para crianças e adolescente também são considerados outros fatores, como: nos comportamentais, o “modo de carregar os materiais escolares” e a “forma de utilizar a mochila”; no hereditário, a “presença de dor nos pais”; e nos socioeconômicos, a “escolaridade do pai” e a “escolaridade da mãe”. Os autores ainda indicam fatores demográficos como a idade e o tipo de escola, mas esses, embora estejam presentes em algum aspecto do sistema de medição, não configuram itens do questionário, não passaram por nenhum método de avaliação em nenhum estudo e não recebem representação. Parecem, portanto, serem aspectos menos relevantes do sistema de medição assim como massa corporal, estatura e profissão.

Nenhum dos artigos apresenta uma caracterização de dor e seus elementos, sendo esse conceito implícito. Quanto à região da dor, a definição é implícita nas primeiras versões do estudo até a proposta de incluir uma imagem no questionário, tanto no BackPEI-CA (Da Rosa *et al.*, 2022) quanto na última versão do BackPEI-A (Candotti *et al.*, 2023). A definição mais explícita aparece nas perguntas feitas ao painel de especialistas durante a avaliação do BackPEI-CA:

A respeito do novo design gráfico das questões sobre dor nas costas, você considera que ele facilita o entendimento que “dor nas costas” se refere a qualquer região torácica e lombar? (Da Rosa et al., 2022, p. 7)

e

A respeito do design gráfico das questões sobre dor no pescoço, você considera que ele facilita o entendimento que “dor no pescoço” se refere a qualquer região da coluna cervical?(Da Rosa et al., 2022, p. 7)

A faixa etária também é definida a partir dos critérios de inclusão dos indivíduos utilizados na avaliação do sistema de medição. Não é apresentada uma definição explícita do que são fatores de risco ou do conceito de cada domínio dos fatores de risco (comportamentais, socioeconômicos e hereditário). Os elementos que compõem os fatores de risco foram identificados a partir de uma revisão de literatura e experiência dos autores, não sendo apresentado um critério objetivo para determina-los. Ainda assim, compreende-se que esses elementos tenham sua determinação a partir da relação com a dor nas costas e a dor no pescoço.

Uma outra forma apresentada de definição do conceito é a partir da sua avaliação científica, ou seja, o que foi feito comparando com outros sistemas de medição que medem um mesmo conceito (Antoniolli *et al.*, 2015), através da verificação das dimensões do conceito (Gençbaş; Bebiş, 2019) e analisando a relação entre conceitos associados (Gökşen; Kocaman; Yıldırım, 2023).

Considerando apenas os itens relativos aos hábitos posturais no BackPEI, Antoniolli e colaboradores (2015) verificaram que há concordância com um sistema de medição que faz as avaliações dos mesmos hábitos posturais através da análise de vídeos, o sistema LADy. De modo semelhante, Gökşen, Kocaman e Yıldırım (2023) encontram relação entre o BackPEI e o MHQ, um sistema de medição que mede conceitos semelhantes ao BackPEI entre pacientes com hérnia de disco. Esses autores mostraram que há uma proximidade conceitual entre os objetos de medição de ambos os sistemas.

Uma forma de verificar a correspondência dos itens de um questionário às dimensões apresentadas na caracterização é através da Análise Fatorial. A aplicação deste método foi relatada na metodologia de Gençbaş e Bebiş (2019) porém, a análise fatorial não pode ser realizada pois os dados não permitiam sua aplicação. Entretanto esse é um método útil à caracterização como pode ser visto no sistema de medição BABAQ desenvolvido a partir do BackPEI (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020). Este instrumento mede o comportamento de cuidado com as costas entre crianças e possui 5 dimensões, definidas a partir da caracterização: habilidades, autoeficácia, conhecimento, crenças de expectativa e comportamento. Após uma análise fatorial confirmatória é apontada a partir de medidas estatísticas como o CFI, AVE e alfa de Cronbach que cada grupo de itens corresponde a uma das dimensões, e apenas a ela, conforme havia sido caracterizado.

As mudanças mais notáveis na caracterização ao longo dos 10 anos de BackPEI foram a inclusão da dor no pescoço e seus fatores de risco relacionados dentro dos conceitos abrangidos pelo sistema de medição, especialmente os fatores relativos ao uso de dispositivos móveis. A ampliação da caracterização foi realizada intencionalmente por parte da equipe de desenvolvimento. O BackPEI sempre teve como um dos seus principais propósitos ser um sistema de medição abrangente, ou seja, permitir a avaliação de vários conceitos através de um único questionário, o que explica a inclusão de uma região a mais na avaliação da dor. Quanto aos fatores de risco relacionados ao uso de dispositivos móveis, a sua inclusão se justifica pela associação com a dor no pescoço. Entretanto, também merece destaque a alteração dos hábitos posturais da população com a expansão do uso de dispositivos móveis. Assim, hábitos que eram menos frequentes em 2013, como utilizar um celular na cama, passaram a ser mais presentes 10 anos depois. Portanto, uma mudança cultural decorrente do avanço tecnológico determina um novo fator de risco que não merecia atenção anteriormente o que reflete em alterações de itens do questionário e inserção de novas questões.

Um outro aspecto de destaque na mudança dos conceitos dos fatores de risco são as diferenças já apresentadas entre crianças e adolescentes em idade escolar e adultos. A diferença de faixa etária também produz diferenças que se refletem na caracterização. Entre os fatores comportamentais, por exemplo, não é relevante avaliar como os adultos carregam a mochila e, mesmo para a dor, a incapacidade de brincar devido a dor é substituída pela incapacidade para trabalhar.

## Representação

Cada aspecto da caracterização corresponde a um item do questionário. As medidas da intensidade da dor nas costas e da dor no pescoço são expressas em centímetros, variando entre 0cm e 10cm. Deste modo, 0cm corresponde a nenhuma dor enquanto uma dor de 10cm é a pior que o indivíduo pode imaginar.

Os demais aspectos têm suas representações possíveis colocadas como alternativas no item correspondente e aquele que responde ao questionário deve escolher uma dessas alternativas. Alguns itens possuem como representações “sim” e “não”: prática de atividade física; prática de esporte competitivo; e envolvimento em acidente de carro. De modo similar os itens relativos a ler e/ou estudar (ou usar dispositivos móveis) na cama possuem as representações “sim”, “não” e “às vezes. Todos os demais itens do questionário possuem como uma de suas alternativas o “não sei responder” ou equivalente (e.g. “depende do dia” e “outro modo”). Alguns destes apresentam nas demais alternativas representações relativas a frequências (Figura 3): frequência de atividade física; tempo assistindo TV; tempo utilizando computador; tempo utilizando dispositivos móveis; tempo de sono; escolaridade da mãe; escolaridade do pai; frequência de dor nas costas; frequência de dor no pescoço. Os demais itens possuem diferentes categorias como opções de representação: todos os itens relativos aos hábitos posturais, à presença de dor (nas costas, no pescoço e nos pais), à incapacidade devido à dor; e o modo de carregar os materiais escolares.



Figura 3 - Exemplo de item com frequências como alternativa

Outra forma de se representar os conceitos do BackPEI é atribuir uma das alternativas como a adequada e as demais como inadequadas, fornecendo uma representação dicotômica. Essa alternativa é utilizada por Antoniolli e colaboradores (2015) para avaliar a concordância do BackPEI com o LADy e, embora os autores tenham utilizado apenas os hábitos posturais em seu estudo, é facilmente adaptável aos demais fatores de risco.

Por fim, a versão original do BackPEI em português, sua tradução para o inglês (Noll *et al.*, 2013a), a versão em turco (Gençbaş; Bebiş, 2019), a versão em espanhol (Miñana-Signes *et al.*, 2021) e a primeira versão do BackPEI-A (Candotti *et al.*, 2018) possuem um sistema de pontuação que atribui uma representação para o conjunto dos fatores de risco. É uma escala discreta na qual quanto maior o valor, menor a exposição a fatores de risco. Essa escala vai de 0 a 10 nos sistemas de medição para adultos (Candotti *et al.*, 2018) e de 0 a 16 naqueles para crianças e adolescentes em idade escolar (Pivotto *et al.*, 2018). Entretanto, não foi feita publicação sobre a atualização desse sistema de pontuação para as versões recentes do BackPEI-A e BackPEI-CA, que consideram a dor no pescoço e o uso de dispositivos móveis (Candotti *et al.*, 2023; Da Rosa *et al.*, 2022).

Quanto aos tipos de escalas utilizados na representação do BackPEI, não é feita nenhuma afirmação explícita. A exceção é a medição da intensidade de dor nas costas. Durante o desenvolvimento do BackPEI original, a primeira versão do questionário foi avaliada por um painel de especialistas que indicaram a alteração da medição da dor em centímetros para uma escala numérica discreta: “os especialistas sugeriram que uma escala numérica fosse incluída, o que transformaria a escala visual analógica em uma escala numérica de dor” (Noll *et al.*, 2013a, p. 569). Nota-se, portanto, uma intenção de que a intensidade da dor seja representada por uma escala contínua.

Através das análises estatísticas aplicadas é possível notar que os autores consideram essa escala como, ao menos, intervalar. Isso é percebido, pois, nas avaliações de confiabilidade teste-reteste são aplicadas análises que não são adequadas a variáveis nominais ou ordinais como o coeficiente de correlação intraclasse e o erro padrão de estimativa. Para as demais representações é possível notar que são tratadas como nominais pelos autores a medida que nas avaliações de confiabilidade teste-reteste são realizadas análises como o percentual de concordância e o coeficiente de kappa, destinados a esse tipo de escala (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a).

Sobre as pontuações que agregam os fatores de risco em uma escala inversa ao risco de dor nas costas não é feita nenhuma alegação e seu valor é utilizado em análises estatísticas apenas na versão em turco entre as fontes analisadas. Gençbas e Bebis (2019) avaliam a confiabilidade teste-reteste baseado nessa pontuação e utilizam o teste dos postos de sinais de Wilcoxon. Gökşen, Kocaman e Yildirim (2023) utilizam do coeficiente de correlação intraclasse para a confiabilidade teste-reteste, o ρ de Spearman para analisar a relação do BackPEI com o MHQ e o teste t independente para comparar os valores superiores e inferiores segundo o BackPEI. Assim é possível ver que os autores assumem que a pontuação total é uma escala ao menos intervalar.

Entre os três elementos da estrutura epistemológica para medições, a representação é o que é menos abordado entre as fontes, inclusive quanto à aplicação de métodos de avaliação científica. Candotti e colaboradoras (2023) a partir das medidas estatísticas obtidos na confiabilidade teste-reteste para a intensidade de dor nas costas e intensidade da dor no pescoço estabeleceu os valores mínimos detectáveis de 2,48cm e 3,10cm respectivamente. Essa medida indica a menor diferença que não pode ser atribuída a um erro de medição (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020)

Gökşen, Kocaman e Yildirim (2023) dividiram a amostra em 2 grupos a partir da pontuação total do BackPEI, um acima do 73º percentil e outro abaixo do 23º. Posteriormente comparou os valores de cada grupo a partir de um teste t. Isso parece uma forma de mostrar uma capacidade discriminatória do BackPEI e demonstrar que há alguma consistência ao atribuir as numerações da escala aos indivíduos.

Além destas, nenhum outro método de avaliação científica com foco principal na representação foi aplicado. Entretanto, toda avaliação realizada a partir dos resultados da medição avalia, em alguma medida, a representação como, por exemplo, o teste-reteste e a validade concorrente. E mesmo a validade de conteúdo, que entre as fontes demonstrou um foco nos procedimentos, permite uma avaliação de algumas representações ao apresentar ao painel de especialistas as alterativas a cada item do questionário.

Uma vez que o BackPEI é um sistema de medição abrangente, é difícil que ele possua uma única representação, dada a variedade de conceitos medidos. Entretanto, há um esforço em sintetizar os valores obtidos por diferentes itens do questionário, especialmente no que se refere aos fatores de risco. Isso pode ser visto em diversas fontes como, por exemplo, no artigo de Pivotto e colaboradoras (2018), que elabora um sistema de pontuação voltado a esse conceito, e no artigo de Candotti e colaboradoras (2018), que já desenvolve a primeira versão do BackPEI-A apresentando um sistema de pontuação. A importância que essa representação tem pode ser vista no estudo de atualização do BackPEI-A, cujas alterações impediram a aplicação do sistema de pontuação anterior, e as autoras afirmam que a “maior limitação deste estudo é a ausência de uma pontuação sobre a possível exposição a fatores associados a dor nas costas e pescoço” (Candotti *et al.*, 2023, p. 163).

Em várias fontes é indicada a intenção de que o BackPEI seja utilizado em pesquisas da ciência da saúde como ensaios clínicos e estudos epidemiológicos e a síntese dos valores em uma pontuação pode contribuir nesse propósito (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a). Também o BackPEI tem como principal propósito avaliar a dor e hábitos posturais em um único sistema de medição com o objetivo de permitir a investigação da associação entre os conceitos, para isso é relevante uma pontuação que abranja todas as posturas como fatores de risco (Candotti *et al.*, 2023; Gençbaş; Bebiş, 2019; Gökşen; Kocaman; Yıldırım, 2023; Noll *et al.*, 2013a).

Assim, vemos que para os fatores de risco existem retrocessos na intenção de estabelecer uma pontuação à medida que há uma atualização da caracterização e, consequentemente, dos procedimentos. Essas atualizações também provocam alterações da representação pois, se cada item do questionário corresponde a uma representação, qualquer inclusão ou remoção de um item também incluem ou removem uma representação.

## Procedimentos

O BackPEI possui como instrumento um questionário auto-aplicável onde cada item é composto por uma questão com as representações como alternativas e o avaliado deve marcar a que considera mais adequada (Figura 3). As exceções são os itens referentes à intensidade de dor, neles há uma escala visual analógica com uma linha de 10cm de comprimento e o avaliado deve identificar nesta linha o ponto equivalente à sua dor com uma marca. Posteriormente um avaliador mede o comprimento entre a marca e o início da linha sendo este o valor da representação da dor (Figura 4). A opção por esta forma de instrumento se deu por este ser barato, de fácil aplicação e por permitir o auto-relato.

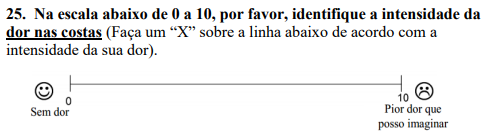


Figura 4 - Escala Visual Analógica da intensidade da dor

Dado que existem diferentes versões do BackPEI também diferem os números de itens do questionário. A questionário mais curto possui 20 itens, caso da primeira versão do BackPEI-A (Candotti *et al.*, 2018), e o mais longo 30, caso do BackPEI-CA (Da Rosa *et al.*, 2022). Responder a essas questões leva de 10 a 20 minutos (Miñana-Signes *et al.*, 2021). De modo geral, esse questionário pode ser aplicado tanto na forma impressa como em formato eletrônico, através de um formulário virtual (Candotti *et al.*, 2023; Da Rosa *et al.*, 2022).

Os hábitos posturais são medidos com o apoio de fotografias conforme o exemplo da Figura 5. Por isso, todas as versões do BackPEI possuem dois questionários, um dedicado ao gênero feminino e outro para o gênero masculino. Com isso pretende-se facilitar a identificação das posturas a partir de imagens correspondentes ao seu gênero e pela diferença de hábitos posturais entre os gêneros. Por exemplo, na Figura 5 as fotografias da direita representam a postura “de perna cruzada” tanto para um gênero quanto para outro, entretanto são posturas diferentes, refletindo hábitos culturais diferentes conforme o gênero. Entretanto, apesar de serem questionários distintos isso não aponta diferenças nem na caracterização, nem na representação. Todas as questões são apresentadas como referentes a um mesmo item e, portanto, a um mesmo conceito. Também os valores obtidos nos questionários para ambos os gêneros são agrupados ao promover análises, demonstrando que é uma diferenciação operacional, sendo assim, de procedimentos.

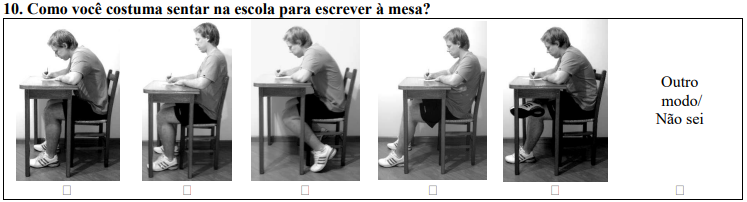


Figura 5 - Medição dos hábitos posturais através de fotografias

Por se tratar de um questionário auto-aplicável a leitura e compreensão das questões e alternativas é um fator operacional importante, o que torna relevante o idioma em que o questionário é escrito. Por isso existem diferentes versões do BackPEI com questionários em diferentes idiomas: português (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Noll *et al.*, 2013a), inglês (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Noll *et al.*, 2013a), turco (Gençbaş; Bebiş, 2019) e espanhol (Miñana-Signes *et al.*, 2021).

A formulação inicial dos itens aconteceu a partir de uma revisão de literatura e da experiência pessoal dos autores. Para os itens referentes aos hábitos posturais, também foi realizada um levantamento de campo através da observação das AVD mais frequentes e as posturas mais comuns nas execuções.

O método de análise dos procedimentos mais empregado foi a confiabilidade teste-reteste (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Gökşen; Kocaman; Yıldırım, 2023; Miñana-Signes *et al.*, 2021; Noll *et al.*, 2013a). Outros métodos empregados foram: a avaliação qualitativa a partir da aplicação do sistema de medição a um grupo de indivíduos (Gençbaş; Bebiş, 2019; Noll *et al.*, 2013a); a comparação dos resultados das medições entre os gêneros (Gökşen; Kocaman; Yıldırım, 2023) e a consulta a um painel de especialistas (validade de conteúdo e adaptação transcultural).

Pela validade de conteúdo foram avaliadas a adequação dos elementos gráficos (fotografias e design gráfico indicando a região referente a dor no pescoço e nas costas); a clareza, facilidade de compreensão e aplicabilidade do instrumento; e a adequação do instrumento ao avaliar dor nas costas, dor no pescoço e seus fatores associados (Candotti *et al.*, 2018, 2023; Da Rosa *et al.*, 2022; Gençbaş; Bebiş, 2019; Noll *et al.*, 2013a). Na adaptação transcultural o painel foi questionado a respeito da semântica, idiomática, experiência e conceito da versão inicial em espanhol após a tradução da versão em português (Miñana-Signes *et al.*, 2021).

Durante os processos de tradução ocorreu algum tipo de avaliação nas traduções do português para o inglês das versões originais do BackPEI e BackPEI-A (Candotti *et al.*, 2018; Noll *et al.*, 2013a) e para o espanhol (Miñana-Signes *et al.*, 2021). Sendo que para a versão em espanhol essa avaliação ocorreu por um método formal, a adaptação transcultural, enquanto que para as versões em inglês essa avaliação parece ter sido sem um método formal.

As alterações de conceitos, como a expansão para adultos e a inclusão da dor no pescoço, demandaram alterações no questionário para incluir os novos itens ou excluir algum antigo. Para além disso algumas alterações nos procedimentos foram realizadas em itens referentes a conceitos que foram mantidos.

No BackPEI-A são utilizadas fotografias diferentes nos itens referentes a hábitos posturais, que passam a utilizar adultos com o objetivo de facilitar a identificação das posturas. Também existem menos opções de posturas para esses itens. Na última versão do BackPEI-CA foi utilizado um novo design gráfico para facilitar a identificação da região referente a dor nas costas e dor no pescoço (Figura 6) (Da Rosa *et al.*, 2022). Esse design foi, posteriormente, implementado no BackPEI-A (Candotti *et al.*, 2023).

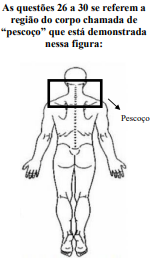
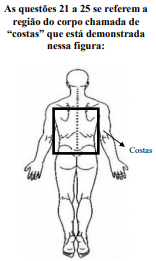


Figura 6 - Design gráfico para apoio aos itens relativos à dor

Durante a coleta de dados para avaliar a confiabilidade teste-reteste do BackPEI-CA, ocorreu a pandemia da Covid-19. Isso acarretou a impossibilidade de aplicação dos questionários impressos. A pesquisa foi interrompida e os pesquisadores adaptaram o questionário para um formulário digital, retomando a coleta de dados de forma virtual. Assim, o sistema de medição foi avaliado com a aplicação nos dois formatos, com a avaliação científica indicando a adequação dos procedimentos segundo a confiabilidade teste-reteste (Da Rosa *et al.*, 2022). Desta forma, há uma alteração dos procedimentos relacionados ao formato de aplicação do questionário. Posteriormente Candotti e colaboradores (2023) avaliaram a confiabilidade teste-reteste do questionário digital, também considerando o procedimento adequado para adultos.

# DISCUSSÃO

Avaliação científica dos elementos

A partir dos resultados percebemos uma valorização do que chamamos de avaliação científica dos elementos. Ou seja, verificar os elementos da estrutura epistemológica (caracterização, representação e procedimentos) utilizando métodos sistemáticos para a coleta de dados, análise e avaliação.

Esse tipo de avaliação não é uma necessidade da medição em si, mas parece relacionado as heurísticas próprias da ciência da saúde, e, portanto, da clinimetria. Por exemplo, entre as formas de se realizar a caracterização temos a definição operacional, quando a caracterização é feita a partir dos próprios procedimentos do sistema de medição (por exemplo, o IMC). Outro modo é fazer a caracterização de forma implícita, através da relação do conceito com outros conceitos (como ocorre com a síndrome metabólica e a fragilidade). E também pode ser feita de modo explícito. (Bradburn; Cartwright; Fuller, 2016; Cartwright; Montuschi, 2018). Mesmo com a ciência produzindo seus conceitos a partir de métodos padronizados de coleta de dados, análise e avaliação isso não significa que esses conceitos, ao serem aplicados em um sistema de medição precisem, necessariamente, passar por um processo de avaliação científica da caracterização, por exemplo. O mesmo vale para a representação e procedimentos. Ainda assim, todos os elementos passaram por alguma avaliação deste tipo por alguma fonte.

Por exemplo, hábitos posturais é um conceito importante para o BackPEI, entretanto não é objetivamente caracterizado por nenhuma das fontes. Porém, é verificada a concordância da medição dos hábitos posturais com o sistema LADy (Antoniolli *et al.*, 2015). Desta forma, parece ser importante para a clinimetria, apresentar métodos para essa avaliação científica da caracterização, representação e procedimentos. Para isso é importante saber a qual elemento cada um dos métodos empregados contempla.

Nessa revisão encontrou-se uma diversidade de termos se referindo a um mesmo método de análise o que dificulta a compreensão da metodologia dos estudos e do propósito de cada método. Este problema também foi identificado em outra revisão de escopo sobre métodos de avaliação dinâmica do ombro que encontrou também o uso de um mesmo termo se referindo a metodologias diferentes (Beraldo; Silva; Candotti, 2022). Os autores indicaram que o uso de guias metodológicos voltados para a clinimetria poderiam auxiliar nesse sentido. Entre as fontes dessa revisão apenas uma relatou utilizar um guia deste tipo. Candotti et al. (2023) informaram seguir o orientado pela iniciativa COSMIN (Mokkink, Lidwine B *et al.*, 2010). Entretanto, a COSMIN identifica parte da metodologia empregada pelas autoras como “erro de medição teste-reteste” que as autoras não reconhecem como uma metodologia específica além de não empregar o termo. E se Candotti e colaboradoras (2023) tivessem adotado o estabelecido pela COSMIN a confusão se agravaria, pois seria a única fonte a empregar o termo “erro de medição teste-reteste” e a única a considera-la como uma metodologia distinta da “confiabilidade teste-reteste”.

Apesar disso, ao indicar uma referência, Candotti e colaboradoras (2023) permitem um esclarecimento sobre o propósito dos métodos empregados. Por exemplo, a COSMIN definia confiabilidade como

o grau no qual uma medição está livre de erros de medição [e] a medida em que as pontuações de pacientes que não mudaram permanecem as mesmas em repetidas medições sob várias condições: por exemplo, [...] ao longo do tempo (teste-reteste) (Mokkink, L B et al., 2010, p. 743).

A partir dessa definição é possível estabelecer que os métodos empregados na confiabilidade teste-reteste estão mais preocupados em avaliar a qualidade operacional do sistema de medição, entendendo que estão voltados aos procedimentos e não a caracterização ou a representação.

Uma forma de evitar essa confusão decorrente da diversidade de termos, metodologias e conceitos seria que os autores apresentassem definições para os termos empregados na metodologia ou fizessem referências a obras que trouxessem essas definições. Permitindo saber exatamente o que o estudo está fazendo e o que pretende com isso.

Eventualmente um mesmo método, ou elementos deles, pode ser aplicado com objetivos distintos. Por exemplo, o índice de kappa foi amplamente utilizado para avaliar os procedimentos através da confiabilidade teste-reteste. Entretanto, essa mesma análise estatística foi aplicada para avaliar a caracterização dos hábitos posturais através da concordância com o sistema LADy.

Outro caso é a validade de conteúdo, que demonstrou uma relação maior com os procedimentos nas pesquisas clinimétricas do BackPEI, mas, para o BABAQ esteve mais relacionada com a caracterização. Essa diferença se deve as perguntas realizadas ao painel de especialistas. Ao avaliar o BackPEI o painel era inquirido sobre aspectos operacionais como a clareza das perguntas e a adequação do uso de fotografias para avaliar hábitos posturais. Já o painel que avaliou o BABAQ foi questionado acerca da necessidade e relevância dos itens, visando reunir no questionário apenas itens que contribuíssem para a medição do comportamento de cuidado com as costas.

Isso não significa que através da validade de conteúdo se avaliou apenas os procedimentos para o BackPEI e apenas a caracterização para o BABAQ, pois os especialistas poderiam apontar a irrelevância de algum item do BackPEI ou a ausência de um aspecto essencial para o questionário deste sistema de medição. Os métodos não foram desenvolvidos tendo em vista a estrutura epistemológica para medições, os pesquisadores buscam avaliar da forma mais ampla possíveis o sistema de medição desenvolvido e as vezes se avalia mais de um elemento de uma vez. Mas isso demostra que a validade de conteúdo, por exemplo, pode avaliar elementos distintos a depender do que se pergunta ao painel de especialistas.

E o que é melhor? Avaliar os procedimentos através da validade de conteúdo ou a caracterização (ou mesmo a representação)?

Entre as fontes dessa revisão se identificou uma dedicação maior aos procedimentos, procedido pela caracterização com a representação no final. Essa dedicação se deu tanto ao explicitar os elementos operacionais nos artigos quanto na avaliação científica realizada. A avaliação dos procedimentos através da confiabilidade teste-reteste é bastante ampla e envolve todos os procedimentos e a validade de conteúdo é importante inclusive para o desenvolvimento de questionários possibilitando o aprimoramento de processos como a substituição de palavras e sugestões de materiais de apoio para a compreensão de perguntas (como as fotografias do BackPEI). Entretanto há uma lacuna sobre como avaliar a caracterização e a representação nas pesquisas clinimétricas do BackPEI.

Na clinimetria, a principal forma de avaliar a caracterização de um sistema de medição é através da validade concorrente. Os mesmos conceitos são medidos nos mesmos objetos de medição por dois sistemas diferentes, o que está sendo avaliado e um sistema que os cientistas confiam que mede aquele conceito. Caso os resultados sejam semelhantes é possível afirmar que ambos os sistemas medem o mesmo conceito. É o mesmo método que foi aplicado na avaliação da concordância do BackPEI com o sistema LADy (Antoniolli *et al.*, 2015). Porém, como o LADy não é considerado o padrão-ouro na medição dos hábitos posturais dificilmente os autores chamariam este método de validade concorrente.

Entretanto, para alguns sistemas de medição não existe a possibilidade de comparação com um padrão-ouro. É o caso de sistemas de medição baseados no auto-relato, como o BackPEI, em que a única forma possível de avaliar a validade concorrente é ao avaliar a versão curta de um questionário em comparação com a versão longa. Assim, a única forma possível de um sistema de medição baseado em auto-relato passar por uma análise pela validade de conteúdo é quando ele, por uma alteração nos procedimentos, apresenta uma vantagem operacional em relação a um sistema de medição já estabelecido destinado ao mesmo conceito.

Uma alternativa é a validade estrutural, método utilizado por Akbari-Chehrehbargh, Tavafian e Montazeri (2020). Por esse método os itens são agrupados conforme sua semelhança segundo testes estatísticos. Esse agrupamento pode comprovar a caracterização na medida que reflete as previsões teóricas feitas posteriormente. Ou pode avançar a caracterização na medida que os autores buscam interpretar os conceitos correspondentes aos conjuntos de itens de um questionário.

Outra opção de avaliação científica da caracterização é através da associação com conceitos que não são o medido pelo sistema de medição. Isso é o que Gökşen, Kocaman e Yildirim (2023) fizeram ao avaliar a correlação do BackPEI com o MHQ, estes não medem os mesmos conceitos, mas possuem alguma semelhança. Ao medir a correlação entre os sistemas espera-se uma correlação moderada e isso é um indicativo da correspondência da medição do sistema com a caracterização do conceito. Esse método, assim como a exploração da relação do conceito medido com outros conceitos e a validade estrutural são boas formas de apontar a caracterização de um sistema de medição. Nesse sentido também é útil que os painéis de especialistas sejam apresentados à caracterização do conceito medido e questionados sobre a correspondência deste com o sistema de medição. Todas essas possíveis soluções apontam para uma mesma necessidade, um maior esforço por parte dos pesquisadores da clinimetria nos aspectos teóricos da caracterização. Por exemplo, trazendo definições explícitas, delimitando as relações que o conceito possui com outros conceitos ou apresentando como os itens de um questionário refletem a caracterização do conceito medido.

Como pode ser visto existe uma certa correspondência entre os métodos de validade e caracterização. Enquanto que métodos de confiabilidade possuem essa relação com os procedimentos. Assim, uma estratégia comum na clinimetria, em avaliar a validade e a confiabilidade de um sistema de medição é uma forma de buscar avaliar tanto a caracterização quanto os procedimentos de um sistema de medição. Entretanto essa não é uma regra, como demonstramos a validade de conteúdo pode ser conduzida de modo a avaliar os procedimentos. As avaliações qualitativas da compreensão dos questionários estão avaliando os procedimentos e foram chamadas de validade de face (Akbari-Chehrehbargh; Tavafian; Montazeri, 2020). Outro exemplo é a avaliação dos itens de um instrumento através do alfa de Cronbach. Entre as fontes essa análise estatística foi considerada parte da confiabilidade. Entretando, é uma medida relacionada com a validade estrutural que demonstra o quanto um conjunto de itens estão correlacionados entre si, indicando a consistência de um instrumento justamente porque essa correlação entre os itens reflete uma correlação com um elemento subjacente a todos eles, o conceito. Assim, embora o alfa de Cronbach, e outras medidas de consistência interna, sejam consideradas medidas de confiabilidade, elas estão mais relacionadas com a caracterização de um conceito em um sistema de medição.

A representação é o elemento da estrutura epistemológica menos abordado entre as fontes. Em geral as escalas são arbitrárias não é estabelecida uma correspondência entre os valores e os conceitos medidos, não é apresentado qual o tipo da escala e não há uma avaliação científica da representação. Uma melhor caracterização do conceito já ajuda nisso, pois esclarece o que o compõe e permite estabelecer correspondências entre os valores da representação e características da conceito.

Por exemplo, a dor é uma percepção. Segundo o BackPEI uma dor com intensidade de 0cm corresponde a nenhuma dor, portanto a uma percepção acessível a qualquer indivíduo que não sentiu dor em algum momento. Entretanto, o limite superior da escala não é tão facilmente atribuível. Pois o que seria o máximo de dor? Existe? A medição proposta atribui o valor de 10cm para esse limite superior e estabelece que ele como a pior dor que o indivíduo sob medição pode imaginar. Assim, dois indivíduos que sintam uma dor com a mesma intensidade podem atribuir valores diferentes na escala se forem diferentes na capacidade de imaginar dores mais intensas que aquela. Uma caracterização da dor poderia esclarecer porque ela possui um limite inferior e um superior e a relação da dor com a capacidade de imaginar a dor. Também poderia gerar suposições sobre o comportamento da escala. Por exemplo se a escala se baseia na imaginação a presença de dor nos pais gera a capacidade de imaginar dores mais intensas, e, portanto, resulta em dores com valores.

Outra contribuição dessa caracterização mais detalhada seria demonstrar melhor a correspondência entre o comprimento da escala e a percepção de dor, isso é importante para demonstrar o tipo de escala. Pois não é porque o comprimento em cm é uma escala racional que a sua utilização para medir a dor gera também uma escala racional para a intensidade da dor. Por exemplo, ainda que 4cm seja metade de um comprimento de 8cm, entretanto não parece óbvio que uma dor de 4cm seja metade de uma dor de 8cm. Na medição é comum que se utilize um fenômeno para medir outro. Por exemplo, termômetros do tipo termopar mede a temperatura a partir da diferença de pontencial. Entretanto a relação entre a diferença de potencial não é linear e os valores precisam ser convertidos a partir de uma equação polinomial de 5º a 9º a depender do termopar. E, mesmo a diferença de potencial medida em volts seja uma escala racional, sua conversão para ºC gera uma escala intervalar. Assim, não é porque uma escala racional é utilizada para medir um efeito relacionado com o conceito sendo medido que os valores dessa escala racional correspondem exatamente às magnitudes do conceito sendo medido.

Determinar o tipo de escala é muito importante pois é assim que vão ser determinados quais os testes estatísticos serão possíveis nas pesquisas que aplicarão o sistema de medição. Variáveis nominais e ordinais não permitem a aplicação de testes paramétricos que são mais valorizados dentro da ciência da saúde e permitem a determinação do poder, tamanho da amostra e tamanho de efeito (ou ao menos facilitam essa determinação ou a fazem de uma forma mais inteligível para os pesquisadores). Nessa revisão nós apresentamos como as fontes encararam a representação, o que, em geral, foi demonstrado através das análises estatísticas empregadas. Assim, as representações da intensidade de dor são sempre tratadas como racionais ou intervalares (pois são avaliadas através do ICC), entretanto, como já apresentado, isso não está evidente.

O mesmo vale para a pontuação do BackPEI que também é tratada como uma variável intervalar ou racional (pois também foi avaliada através do ICC). Entretanto, por apresentar pesos iguais a todos os fatores de risco ela só pode corresponder a uma escala intervalar se o risco de dor nas costas ou no pescoço for exatamente o mesmo para cada fator e se a inter-relação entre todos os fatores for igual, o que as revisões de literatura utilizadas no desenvolvimento do BackPEI demonstram que não é o caso.

Só que pra demonstrar essas coisas é preciso usar (problema da medição nômica)

# CONCLUSÃO

# rEFERÊNCIAS