

## Construção de uma escala para avaliar a qualidade metodológica de revisões sistemáticas

Construction of a scale to assess the methodological quality of systematic reviews

Angelo Brandelli Costa <sup>1</sup>  
Ana Paula Couto Zoltowski <sup>1</sup>  
Silvia Helena Koller <sup>1</sup>  
Marco Antônio Pereira Teixeira <sup>1</sup>

**Abstract** *The aim of this study was to describe the process of creation and validation of methodology to assess the methodological quality of systematic reviews. From the expansion of the scope and addition of new items to the Assessment of Multiple Systematic Reviews (AMSTAR) instrument evidence of validity and reliability of the new methodology was sought. Firstly, the original scale was translated, adapted, expanded and evaluated by a panel of five experts. Then it was tested by evaluating a random sample of five systematic reviews by three judges with no formal training in the scale. Finally, the consolidated scale was assessed in a new panel by two judges. The panel of experts reached a consensus regarding the validity of the proposed changes. With respect to the scale total and per item score, the average agreement between judges was considered satisfactory in the consolidated scale. From the processes of adaptation and validation performed, the methodology in its final version consisted of 14 items that taken together indicate the methodological quality of systematic reviews from different areas of knowledge.*

**Key words** *Systematic review, Methodology, Quality*

**Resumo** *O objetivo deste estudo foi descrever o processo de construção e de validação de uma metodologia para avaliação da qualidade de escalas de revisões sistemáticas. A partir da ampliação do escopo e de acréscimos de novos itens ao instrumento Assessment of Multiple Systematic Reviews (AMSTAR), buscaram-se evidências de validade e confiabilidade da nova metodologia. Em um primeiro momento, a escala original foi traduzida, adaptada, ampliada e avaliada por um painel de cinco especialistas. Em seguida, foi utilizada para avaliar uma amostra aleatória de cinco revisões sistemáticas por três juízes sem treinamento formal na escala. Por fim, a escala consolidada foi avaliada em uma nova rodada por 2 juízes. O painel de especialistas entrou em consenso em relação à validade das alterações propostas. Em relação à escala consolidada, a concordância média entre os juízes por itens e no score total foi considerada satisfatória. A partir dos processos de adaptação e validação realizados, a metodologia em sua versão final foi composta por 14 itens que conjuntamente indicam a qualidade metodológica de revisões sistemáticas de campos do conhecimento diversos.*

**Palavras-chave** *Revisão sistemática, Metodologia, Qualidade*

<sup>1</sup> Instituto de Psicologia,  
Universidade Federal do Rio  
Grande do Sul. R. Ramiro  
Barcelos 2600, Santa Cecília.  
90035-003 Porto Alegre  
RS Brasil.  
brandelli.costa@ufrgs.br

## Introdução

A revisão de artigos é uma das práticas mais recorrentes na literatura científica. Atualmente, destaca-se o uso da técnica de revisão sistemática (RS) por ser um processo formal, organizado e controlado para avaliação e síntese de estudos em diversas áreas do conhecimento. No entanto, a ideia contemporânea de uma revisão de pesquisas feita de maneira sistemática surgiu apenas no final da década de 1970, a partir do conceito de meta-análise<sup>1</sup>. Apesar do termo meta-análise ser frequentemente confundido como sinônimo de RS, os dois termos possuem sentidos distintos<sup>2</sup>. Meta-análise refere-se ao procedimento estatístico de tratamento dos dados de diversos estudos com o objetivo de agrupá-los, enquanto que RS refere-se ao processo de reunião, avaliação crítica e sintética dos resultados de múltiplos estudos, podendo ou não incluir meta-análise<sup>3</sup>.

As RS são ferramentas importantes para subsidiar decisões de políticas e intervenções em saúde, uma vez que fornecem informações com um viés de publicação mais controlado. O viés de publicação pode ser entendido pela tendência de estudos com resultados positivos (ou significativos) serem mais aceitos para publicação em detrimento daqueles com resultados negativos (ou não esperados)<sup>4</sup>. Assim, o pesquisador, ao realizar uma busca não sistemática, acaba supervalorizando a seleção de artigos com resultados positivos, visto que aqueles com resultados negativos costumam ser mais difíceis de serem encontrados. Buscando minimizar o viés de publicação, a RS caracteriza-se pela aplicação de estratégias de busca, análise crítica e síntese da literatura de forma organizada. Dessa forma, revisar sistematicamente é um processo muito mais amplo do que apenas descrever os resultados encontrados, pois implica a adoção de uma análise reflexiva dos dados, sejam eles quanti ou qualitativos<sup>5,6</sup>.

As RS podem apresentar limitações metodológicas importantes, tanto no que se refere à estratégia de busca quanto à análise crítica da qualidade dos estudos incluídos<sup>7</sup>. Soma-se a isso a constatação de que, em alguns campos do conhecimento, as RS parecem ainda não apresentar um aprimoramento significativo de qualidade que acompanhe a crescente quantidade de publicações<sup>8-10</sup>. Portanto, mesmo sendo uma técnica já estabelecida, uma RS, para ter sua validade assegurada, necessita garantir uma boa qualidade metodológica, pois isso aumenta a probabilidade da apresentação de resultados não enviesados,

além de ser um pré-requisito para interpretações e aplicações válidas<sup>11</sup>.

Observa-se a presença de diferentes instrumentos, *checklists* e *guidelines* para identificar, analisar e sintetizar os dados, ocasionando uma variabilidade na maneira de se conduzir uma RS<sup>12</sup>. Por exemplo, existem *checklists* desenvolvidos especificamente para avaliar a estrutura formal dos artigos de RS. Esse é o caso do Qualitative Assessment and Review Instrument (QARI)<sup>13</sup> e do Meta-Analysis of Statistics Assessment and Review Instrument (MasTARI), ambos do Joanna Briggs Institut. Para além dos aspectos formais, a qualidade metodológica de uma RS representa o quão bem ela foi conduzida, isto é, a sua validade interna. É importante ressaltar que a qualidade metodológica de uma RS não deve ser confundida com a dos estudos revisados, embora esteja relacionada com a avaliação e a descrição cuidadosa destes.

Na literatura internacional, pode-se identificar mais de 24 instrumentos para avaliar especificamente a qualidade metodológica de RS<sup>14</sup>. Um dos poucos que foi validado e desenvolvido de forma rigorosa denomina-se Assessment of Multiple Systematic Reviews (AMSTAR)<sup>15</sup>. A construção e a validação do AMSTAR são apresentadas, respectivamente, em dois estudos descritos a seguir.

O AMSTAR foi construído a partir da combinação das seguintes fontes de informação: 1) o Overview Quality Assessment Questionnaire (OQAQ)<sup>16</sup> que contém 10 itens; 2) um checklist criado por Sacks et al.<sup>17</sup>, contendo 24 itens; e 3) três itens adicionais que os autores julgaram ser de importância metodológica ao instrumento (restrição do idioma de publicação como critério de seleção dos estudos, o viés de publicação e a origem da publicação, se tese, dissertação, artigos, resumos, etc.). A versão inicial da ferramenta foi composta de 37 itens, sendo aplicada na avaliação de 99 artigos provenientes de bases de dados de revisões e meta-análises e 52 publicações impressas provenientes da base Cochrane de RS. Para cada revisão, dois juízes independentes avaliaram a sua qualidade metodológica. Posteriormente, realizou-se uma análise fatorial exploratória com todos os itens, e apenas aqueles que carregaram score acima de 0,50 foram mantidos, o que possibilitou um instrumento mais curto, com 29 que mediam onze componentes principais<sup>15</sup>.

Após a etapa das análises estatísticas, um grupo de pesquisadores e pareceristas sem treinamento no instrumento discutiram os itens consi-

derados mais apropriados para cada componente para a inclusão na versão final do instrumento. O instrumento final foi formado por onze itens, com uma grade de respostas composta por 'Sim', 'Não', 'Não é possível responder' ou 'Não se aplica'. O item que foi criado na primeira versão do instrumento a fim de avaliar restrição do idioma de publicação, mesmo sendo identificado na análise fatorial, acabou por não fazer parte do instrumento na versão final, devido a inconsistências empíricas apontadas pela literatura sobre esse tópico<sup>15</sup>.

Após o processo de construção do instrumento, os autores buscaram mensurar a concordância, a fidedignidade, a validade de construto e a aplicabilidade do AMSTAR<sup>11</sup>. A partir de um banco de 151 RS, 30 artigos (20%) foram selecionados aleatoriamente a fim de serem avaliados. Dois juízes (um sem treinamento formal no instrumento) aplicaram o AMSTAR e mais dois outros instrumentos de avaliação da qualidade metodológica – o OQAQ 16 e o *checklist* desenvolvido por Sacks et al.<sup>17</sup>. De forma geral, o AMSTAR demonstrou ter boa concordância, confiabilidade, validade de construto e aplicabilidade, mostrando-se similar ao OQAQ em termos de concordância e confiabilidade e melhor que o *checklist* de Sacks. Por fim, sugestões foram dadas para o contínuo processo de validação do instrumento, a fim de se replicar seus resultados iniciais com um grupo mais variado de revisores e de artigos avaliados<sup>11</sup>.

Embora os autores afirmem que o AMSTAR tenha sido desenvolvido e adaptado para avaliar RS de todas as áreas, percebe-se que o instrumento peca na avaliação de estudos qualitativos, comuns no campo da saúde coletiva, pois não compreende métodos de análise utilizados para pesquisas desse tipo. Os critérios-base dos itens do instrumento apresentam uma preponderância pela síntese dos dados através de análises quantitativas, o que não é aplicável a todos os contextos, como o da saúde coletiva. Além disso, o AMSTAR possui algumas fragilidades psicométricas importantes, como avaliação de mais de um critério em um mesmo item, e ambiguidades nas suas descrições e sistema de pontuação, o que pode provocar inconsistências na utilização<sup>18</sup>. Dessa forma, constatando a ausência de instrumentos com esta finalidade no contexto brasileiro, o objetivo deste estudo foi descrever o processo de construção e de validação de uma nova metodologia para avaliação da qualidade metodológica de RS, tendo como ponto de partida o AMSTAR.

## Método

### Procedimentos

O procedimento para construção do instrumento foi baseado em Borsa et al.<sup>19</sup> de acordo com as seguintes etapas: 1) Tradução, equivalência contextual e revisão por comitê de especialistas; 2) estudo piloto; e 3) estudo de validação. Não foi realizada a etapa de retrotradução, uma vez que se trata da ampliação de um instrumento, cuja versão final difere substancialmente da original.

### Tradução, equivalência contextual, ampliação e revisão por comitê de especialistas

Em um primeiro momento, a escala foi traduzida do inglês para o português, de maneira independente, por dois tradutores. As traduções foram comparadas e sintetizadas por um terceiro tradutor independente. Os três tradutores são proficientes nas línguas portuguesa e inglesa, além de serem familiarizados com a temática. Em seguida, os acréscimos, as supressões e a equivalência dos itens que compunham o instrumento original foram analisados a partir de um comitê de cinco juízes especialistas na publicação de revisões sistemáticas e estudos na área de psicometria. O comitê ponderou a pertinência semântica, idiomática e conceitual de cada item, além da sua consigna e, quando necessárias, redações alternativas foram sugeridas. A pertinência dos itens foi classificada entre “pertinente”, “pertinente com alterações” e “não pertinente”. A versão traduzida, modificada e a ampliada pode ser encontrada no Quadro 1.

Os critérios de pontuação foram alterados para apenas “Sim” e “Não”. Optou-se pela remoção das opções “Não é possível responder” e “Não se aplica”, pois poderiam levar à heterogeneidade nas avaliações já que concernem fortemente ao arbítrio do avaliador. No entanto, ressalta-se que a nova opção de resposta não fugiu da estrutura de pontuação original, pois nesta, mesmo com quatro opções, apenas a resposta “Sim” era pontuada. Salienta-se ainda que a mesma opção simplificada já havia sido exposta com sucesso na literatura recente<sup>8</sup>. Informações sobre possíveis exceções foram incluídas nas consignas dos itens pertinentes.

**Quadro 1.** Itens e critérios do instrumento original e da adaptação.

Itens do instrumento original	Crítérios do instrumento original	Itens adaptados	Crítérios dos itens adaptados
1. Was an “a priori” design provided?	<i>The research question and inclusion criteria should be established before the conduct of the review.</i>	1. O objetivo de pesquisa está claramente descrito?	<i>O objetivo de pesquisa deve estar claramente descrito antes do começo da revisão.</i>
		2. Os elementos essenciais que compõem a estratégia de busca são descritos?	<i>As palavras-chave e os operadores booleanos (e; ou; não; etc.), quando utilizados, devem ser indicados. Recortes temporais na busca, se realizados, devem ser descritos de forma clara. A data em que foi realizada a busca deve ser descrita.</i>
		4. Os critérios de inclusão e exclusão utilizados na seleção dos estudos foram descritos?	<i>Crítérios metodológicos, como delineamento utilizado (e.g., experimentos, levantamentos, estudos de caso); tipo de instrumento utilizado (e.g., entrevistas, grupos focais, testagem, etc.); e análise dos dados (e.g., análise de conteúdo, fenomenologia, testes estatísticos, etc.) podem ser descritos como critérios para seleção dos estudos, entre outros.</i>
2. Was there duplicate study selection and data extraction?	<i>There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.</i>	6. No mínimo dois juízes realizaram a busca e a seleção das publicações?	<i>No mínimo dois juízes independentes realizaram a busca e selecionaram as publicações com base nos critérios de inclusão e exclusão. Estratégias de consenso, em caso de discordância, devem ser informadas.</i>
		8. No mínimo dois juízes realizaram a extração de dados?	<i>No mínimo, dois juízes independentes realizaram a extração de dados nos artigos selecionados. Estratégias de consenso, em caso de discordância, devem ser informadas.</i>
3. Was a comprehensive literature search performed?	<i>At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.</i>	5. Uma busca abrangente foi realizada?	<i>No mínimo, duas bases de dados eletrônicas devem ser utilizadas (Psycinfo, Scielo, Medline, etc.) e descritas. A pesquisa pode ser complementada por outros materiais, tais como revisões, livros-texto, textos técnicos, buscando nas referências dos estudos encontrados e consultando especialistas no campo específico de estudo.</i>

continua

Quadro 1. continuação

Itens do instrumento original	Critérios do instrumento original	Itens adaptados	Critérios dos itens adaptados
4. Was the status of publication (i.e., grey literature) used as an inclusion criterion?	<i>The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.</i>	3. A origem da publicação foi utilizada como critério de busca?	<i>Os autores devem descrever se não incluíram alguma publicação devido a sua origem (tese, dissertação, resumo de congresso), idioma, etc.</i>
5. Was a list of studies (included and excluded) provided?	<i>A list of included and excluded studies should be provided.</i>	7. O número de artigos incluídos e excluídos em cada etapa da busca foi informado?	<i>Deve ficar explícito, seja em texto, figura, lista ou gráfico, quantos artigos foram incluídos e excluídos em cada etapa da revisão.</i>
6. Were the characteristics of the included studies provided?	<i>In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.</i>	9. Foram descritas as características dos estudos incluídos?	<i>Devem ser descritos, de acordo com os objetivos do estudo, de forma integrada, seja em uma tabela ou textualmente, as características dos estudos revisados. Por exemplo, idade, raça, sexo, dados socioeconômicos relevantes, delineamentos, técnicas de amostragem, desfechos investigados, etc.</i>
7. Was the scientific quality of the included studies assessed and documented?	<i>'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.</i>	10. A forma de avaliação da qualidade metodológica dos estudos incluídos foi descrita?	<i>A forma de avaliação da qualidade dos estudos revisados deve ser relatada. Por exemplo, pode-se levar em consideração diferentes tamanhos de amostra, tipos de delineamento, poder estatístico dos testes utilizados, tamanhos de efeitos, etc. No caso de estudos qualitativos incluídos na seleção, os autores devem avaliar a análise de dados realizada e se esta se mostra embasada teoricamente, se há uma descrição do procedimento de codificação e/ou análise de dados, incluindo quem codificou e quais foram as unidades de análise, etc.</i>
8. Was the scientific quality of the included studies used appropriately in formulating conclusions?	<i>The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.</i>	11. A qualidade metodológica dos estudos incluídos foi utilizada apropriadamente na formulação das conclusões?	<i>As conclusões da revisão devem levar em consideração as diferentes qualidades metodológicas dos estudos revisados, discutindo as limitações desses estudos, especialmente na formulação de recomendações futuras.</i>

continua

### Estudo Piloto

Para essa etapa, foi utilizada uma amostra aleatória de cinco RS. Os temas das revisões variaram

em todo o espectro da Psicologia. Esta área do conhecimento foi escolhida pela sua proximidade com o campo de atuação dos autores e também por abarcar estudos com métodos qualitativos e

Quadro 1. continuação

Itens do instrumento original	Crítérios do instrumento original	Itens adaptados	Crítérios dos itens adaptados
9. Were the methods used to combine the findings of studies appropriate?	<i>For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chisquared test for homogeneity, I2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).</i>	12. Os métodos utilizados para integrar os resultados dos estudos foram descritos?	<i>Os métodos para integrar os resultados devem ser descritos e apropriados, tanto nas revisões que utilizam análises quantitativas quanto nas que adotam análises qualitativas. Caso o autor queira integrar resultados de estudos quantitativos, testes devem ser realizados para assegurar a possibilidade de combinar os estudos revisados (por exemplo, teste qui-quadrado para homogeneidade). Técnicas de integração devem ser descritas, por exemplo, meta-análises. Em caso de não homogeneidade, deve ser considerada a viabilidade de combinar tais resultados. No caso de revisões que utilizam estudos qualitativos e/ou que se propõem a uma análise qualitativa dos resultados, métodos específicos de análise para a integração devem ser reportados (por exemplo, análise do discurso, análise de conteúdo, grounded theory, meta-síntese). Caso o autor opte por não realizar a integração dos estudos revisados, no caso de uma revisão descritiva, por exemplo, essa opção deve ser descrita.</i>
10. Was the likelihood of publication bias assessed?	<i>An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).</i>	13. O viés de publicação foi considerado?	<i>A possibilidade que os dados possam estar enviesados em virtude de alguma limitação da própria revisão deve ser considerada. Uma análise do viés de publicação pode incluir uma combinação gráfica (por exemplo, funnel plot ou outros testes disponíveis) e/ou testes estatísticos (por exemplo, teste de regressão de Egger). No mínimo, o viés de publicação deve ser textualmente apontado como um dos limitadores da revisão realizada.</i>
11. Was the conflict of interest included?	<i>Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.</i>	14. O conflito de interesses foi descrito?	<i>Fontes potenciais de apoio devem ser claramente apontadas tanto na revisão sistemática quanto nos estudos incluídos. Se não há conflito de interesses, de igual maneira, o autor deve informar ao leitor a ausência de fontes potenciais de conflito.</i>

quantitativos, aproximando-se tanto das ciências humanas e sociais quanto das biológicas e da saúde. Além disso, devido ao instrumento possuir

uma preponderância por análises quantitativas, buscou-se testar a versatilidade da metodologia frente a uma gama heterogênea de estudos.



Três juízes (A, B e C), sem treinamento formal na escala, avaliaram as revisões utilizando a versão adaptada do instrumento. A cada item marcado com “Sim” foi atribuído um ponto e a soma de todos os itens marcados afirmativamente constituiu a pontuação total. Os coeficientes de correlação intraclassa (CCI) foram estimados para avaliar a confiabilidade da pontuação total. Ainda foram estimados os valores do Kappa de Cohen para cada item a fim de analisar a concordância interobservador, sendo que itens que apresentaram valores Kappa abaixo de 0,40 foram revisados novamente e novas redações foram sugeridas. Seguindo os mesmos critérios de Shea et al.<sup>11</sup>, valores de Kappa inferiores a 0 foram classificados como de acordo ao acaso; de 0,01 a 0,20, como concordância leve; de 0,21 a 0,40, como concordância baixa; de 0,41 a 0,60 como concordância moderada; de 0,61 a 0,80, como acordo substancial; e de 0,81 a 0,99, como concordância quase perfeita.

### Estudo de Validade

Por fim, realizou-se uma nova avaliação a partir dos itens consolidados na etapa anterior. Dois juízes, sem treinamento formal na escala, avaliaram 25 RS escolhidas aleatoriamente de artigos do âmbito da Psicologia. Foram utilizados os mesmos critérios e procedimentos de análise do estudo piloto.

### Resultados

A partir do instrumento original, o processo de construção propôs a criação de itens adicionais para evitar ambiguidades e adequar o instrumento a um contexto mais diversificado de estudos (Quadro 1). O item 1 do instrumento original (o desenho da revisão foi apresentado a priori?) foi desmembrado em três novos itens que consideravam todos os critérios elencados no item original: 1) O objetivo de pesquisa está claramente descrito? 2) Os elementos essenciais que compõem a estratégia de busca são descritos? 4) Os critérios de inclusão e exclusão utilizados na seleção dos estudos foram descritos? Além disso, o item 2 do instrumento original (havia duplicação na extração de dados e seleção dos estudos?) foi desmembrado em dois itens, um perguntando a respeito da busca e seleção de estudos (6) e outro, da extração de dados (8). A redação de alguns itens também foi refeita. Por exemplo, o item 7 do instrumento converteu-se em: 10) A forma de

avaliação da qualidade metodológica dos estudos incluídos foi descrita? Uma vez que esse era o objetivo do item em questão. A descrição de alguns itens foi alterada de forma a abranger aspectos metodológicos de revisões que incluam também pesquisas qualitativas. Esse foi o caso dos itens 1, 7 e 9 do instrumento original. Por fim, os itens foram reordenados de forma a acompanhar a ordem lógica da realização de uma RS.

Em seguida, um painel de especialistas discutiu a pertinência dos itens para o objetivo do instrumento. Todos os itens foram julgados pertinentes, sendo que alguns ajustes de redação foram propostos e contemplados. O instrumento final é um questionário de 14 itens para serem respondidos entre “Sim” ou “Não” a partir da presença ou da ausência de cada critério. A versão do instrumento foi testada com cinco RS avaliadas por três juízes que não tinham experiência prévia com o instrumento. O CCI para o escore total foi excelente: 0,925 (95% CI: 0,65 - 0,99)  $p < 0,001$ ; já a concordância interjuízes foi moderada: kappa médio 0,50 (intervalo: -0,07 e 1,0). No entanto, os itens 2 (estratégia de busca), 3 (origem da publicação), 4 (critérios de inclusão e de exclusão), 9 (características dos estudos) e 11 (método para integrar os resultados) tiveram pontuações baixas em - 0,07, - 0,07, 0,2, - 0,07 e - 0,07 respectivamente (Tabela 1). Com o auxílio dos juízes, a análise qualitativa destes itens levou a pequenas modificações em sua redação, a fim de esclarecer o propósito das questões e evitar ambiguidades. A versão final do instrumento pode ser encontrada no Quadro 2.

**Tabela 1.** Concordância interjuízes do estudo piloto.

Itens	Kappa	IC 95%	Concordância
1	1,00	-	Perfeita
2	-0,36	[-0,87, 0,14]	Ao acaso
3	-0,20	[-0,71, 0,31]	Ao acaso
4	0,10	[-0,41, 0,61]	Leve
5	1,0	-	Perfeita
6	0,42	[-0,08, 0,93]	Moderada
7	0,66	[0,15, 1,00]	Substancial
8	1,0	-	Perfeita
9	-0,07	[-0,58, 0,44]	Ao acaso
10	0,17	[-0,34, 0,67]	Leve
11	-0,07	[-0,58, 0,44]	Ao acaso
12	0,66	[0,15, 1,00]	Substancial
13	1,00	-	Perfeita
14	1,00	-	Perfeita

**Quadro 2.** Versão final do instrumento.

<p>As perguntas abaixo se referem a critérios metodológicos minimamente necessários para uma revisão sistemática de qualidade. As perguntas estão acompanhadas de um texto explicativo para auxiliar sua compreensão. Para cada pergunta, marque SIM ou NÃO, de acordo com a presença ou ausência do critério na revisão avaliada. Em caso de dúvida e/ou falta de clareza em relação à presença de algum critério, marque NÃO.</p>	
<p>1) O objetivo de pesquisa está claramente descrito? <i>O objetivo de pesquisa deve estar claramente descrito antes do começo da revisão.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>2) Os elementos essenciais que compõem a estratégia de busca são descritos? <i>As palavras-chave e os operadores booleanos (e; ou; não; etc.), quando utilizados, devem ser indicados. Recortes temporais na busca, se realizados, devem ser descritos de forma clara. A data em que foi realizada a busca deve ser descrita.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>3) A origem da publicação foi utilizada como critério de busca? <i>Os autores devem descrever se não incluíram alguma publicação devido à sua origem (tese, dissertação, resumo de congresso) ou idioma, etc.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>4) Os critérios de inclusão e exclusão utilizados na seleção dos estudos foram descritos? <i>Critérios metodológicos, como delineamento utilizado (e.g., experimentos, levantamentos, estudos de caso); tipo de instrumento utilizado (e.g., entrevistas, grupos focais, testagem, etc.); e análise dos dados (e.g., análise de conteúdo, fenomenologia, testes estatísticos, etc.) podem ser descritos como critérios para seleção dos estudos, entre outros.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>5) Uma busca abrangente foi realizada? <i>No mínimo, duas bases de dados eletrônicas devem ser utilizadas (Psycinfo, Scielo, Medline, etc.) e descritas. As pesquisas podem ser complementadas por outros materiais tais como revisões, livros-texto, textos técnicos, buscando nas referências dos estudos encontrados e consultado especialistas no campo específico de estudo.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>6) No mínimo dois juízes realizaram a busca e a seleção das publicações? <i>No mínimo dois juízes independentes realizaram a busca e selecionaram os estudos com base nos critérios de inclusão e exclusão. Estratégias de consenso em caso de discordância devem ser informadas.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>7) O número de artigos incluídos e excluídos em cada etapa da busca foi informado? <i>Deve ficar explícito, seja em texto, figura, lista ou gráfico, quantos artigos foram incluídos e excluídos em cada etapa da revisão.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>8) No mínimo dois juízes realizaram a extração de dados? <i>No mínimo, dois juízes independentes realizaram a extração de dados nos artigos selecionados. Estratégias de consenso, em caso de discordância, devem ser informadas.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO
<p>9) Foram descritas as características dos estudos incluídos? <i>Devem ser descritos, de acordo com os objetivos do estudo, de forma integrada, seja em uma tabela ou textualmente, as características dos estudos revisados. Por exemplo, idade, raça, sexo, dados socioeconômicos relevantes, delineamentos, técnicas de amostragem, desfechos investigados, etc.</i></p>	<input type="checkbox"/> SIM <input type="checkbox"/> NÃO

continua

Realizou-se uma nova avaliação a partir da versão atualizada do instrumento proveniente do estudo piloto. Dessa forma, dois juízes avaliaram

25 RS escolhidas aleatoriamente. A concordância interjuízes variou de moderada (0,42) à perfeita (1,0). Os itens que foram modificados em função



Quadro 2. continuação

<p>10) A forma de avaliação da qualidade metodológica dos estudos incluídos foi descrita?  <i>A forma de avaliação da qualidade dos estudos revisados estabelecida deve ser relatada. Por exemplo, pode-se levar em consideração diferentes tamanhos de amostra, tipos de delineamento, poder estatístico dos testes utilizados, tamanhos de efeitos, etc. No caso de estudos qualitativos incluídos na seleção, os autores devem avaliar a análise de dados realizada e se esta se mostra embasada teoricamente, se há uma descrição do procedimento de codificação e/ou análise de dados, incluindo quem codificou e quais foram as unidades de análise, etc.</i></p>	<table border="1"> <tr> <td>SIM</td> <td>NÃO</td> </tr> </table>	SIM	NÃO
SIM	NÃO		
<p>11) Os métodos utilizados para integrar os resultados dos estudos foram descritos?  <i>Os métodos para integrar os resultados devem ser descritos e apropriados, tanto nas revisões que utilizam análises quantitativas quanto nas que adotam análises qualitativas. Caso o autor queira integrar resultados de estudos quantitativos, testes devem ser realizados para assegurar a possibilidade de combinar os estudos revisados (por exemplo, teste qui-quadrado para homogeneidade). Técnicas de integração devem ser descritas, por exemplo, meta-análises. Em caso de não homogeneidade, deve ser considerada a viabilidade de combinar tais resultados. No caso de revisões que utilizam estudos qualitativos e/ou que se propõem a uma análise qualitativa dos resultados, métodos específicos de análise para a integração devem ser reportados (por exemplo, análise do discurso, análise de conteúdo, groundedtheory, meta-síntese). Caso o autor opte por não realizar a integração dos estudos revisados, no caso de uma revisão descritiva, por exemplo, essa opção deve ser descrita.</i></p>	<table border="1"> <tr> <td>SIM</td> <td>NÃO</td> </tr> </table>	SIM	NÃO
SIM	NÃO		
<p>12) A qualidade metodológica dos estudos incluídos foi utilizada apropriadamente na formulação das conclusões?  <i>As conclusões da revisão devem levar em consideração as diferentes qualidades metodológicas dos estudos revisados, discutindo as limitações desses estudos, especialmente na formulação de recomendações futuras.</i></p>	<table border="1"> <tr> <td>SIM</td> <td>NÃO</td> </tr> </table>	SIM	NÃO
SIM	NÃO		
<p>13) O viés de publicação foi considerado?  <i>A possibilidade que os dados possam estar enviesados em virtude de alguma limitação da própria revisão deve ser considerada. Uma análise do viés de publicação pode incluir uma combinação gráfica (por exemplo, funnelplot ou outros testes disponíveis) e/ou testes estatísticos (por exemplo, teste de regressão de Egger). No mínimo, o viés de publicação deve ser textualmente apontado como um dos limitadores da revisão realizada.</i></p>	<table border="1"> <tr> <td>SIM</td> <td>NÃO</td> </tr> </table>	SIM	NÃO
SIM	NÃO		
<p>14) O conflito de interesses foi descrito?  <i>Fontes potenciais de apoio devem ser claramente apontadas tanto na revisão sistemática quanto nos estudos incluídos. Se não há conflito de interesses, de igual maneira, o autor deve informar ao leitor a ausência de fontes potenciais de conflito.</i></p>	<table border="1"> <tr> <td>SIM</td> <td>NÃO</td> </tr> </table>	SIM	NÃO
SIM	NÃO		
<p>TOTAL</p>			

das pontuações baixas apresentadas no estudo piloto demonstraram um melhor grau de concordância nesta etapa, sendo que os itens 3, 9 e 11 passaram a apresentar concordância moderada, enquanto o item 2 apresentou concordância substancial.

Destaca-se que os itens que obtiveram maior grau de concordância foram o 14 (conflito de interesses), o 13 (viés de publicação) e o 5 (uma

busca abrangente foi realizada), com valores kappa de 1,0; 0,86 e 0,66, respectivamente. Por outro lado, os que apresentaram menor grau de concordância entre os juízes foram os itens 4 (critérios de inclusão e de exclusão), 7 (número de artigos incluídos e excluídos em cada etapa) e 12 (qualidade metodológica trazida na conclusão), com valores kappa de 0,42; 0,42 e 0,43; respectivamente. A concordância média interjuízes foi

moderada: 0,52 (intervalo: 0,42 e 1,0). Novamente, o CCI para o escore total do instrumento foi excelente: 0,805 (95% CI: 0,56 - 0,91)  $p < 0,001$  (Tabela 2).

## Discussão

A avaliação da qualidade metodológica de RS ainda é uma temática incipiente no contexto brasileiro. A construção, a partir de um instrumento internacionalmente reconhecido, e a validação de uma escala ampliada para este fim foram os objetivos deste estudo. Durante esse processo, foram necessárias modificações nos itens, nas suas descrições e na sua grade de respostas, a fim de deixar o instrumento menos ambíguo, facilitando sua aplicabilidade. Nesse sentido, foram acrescentados três itens à versão original, frutos do desmembramento daqueles já existentes. Buscou-se também adotar uma descrição mais minuciosa, a fim de clarificar o critério avaliativo. Na etapa de validade de construto, todos os itens foram considerados pertinentes pelo comitê de especialistas, mesmo que alguns apontassem a necessidade de alterações, o que sugere que os critérios da escala são úteis também no panorama acadêmico e editorial brasileiro.

Embora os indicadores de concordância entre juízes tenham sido satisfatórios, tanto na pontuação geral do instrumento quanto na análise item-a-item, ressalta-se que quatro itens (2, 3, 4 e 11) obtiveram pontuações classificadas como discordantes entre os avaliadores. Cada um destes itens foi analisado junto com os juízes, buscando-se

investigar os motivos para as discordâncias e definir possíveis modificações. Com isso, no segundo item, alterou-se a sua descrição, detalhando quais eram os elementos essenciais na composição da estratégia de busca, principalmente enfatizando a necessidade de se apresentar o período em que as revisões foram buscadas. Já para o terceiro item, mudou-se a escrita da sua pergunta, salientando-se a importância de se explicitar a inclusão ou não de materiais provenientes de fontes de divulgação diferentes (resumos de congresso, dissertações, teses, artigos, etc.). No item 4 foi realizada uma modificação na escrita da sua descrição, indicando que os critérios citados ali tinham a finalidade apenas de exemplificar a gama de critérios de inclusão e de exclusão possíveis, e não de indicar quais critérios deveriam estar presentes na revisão avaliada. Por fim, o item 11 demandou maior detalhamento em sua descrição, principalmente no caso de revisões que integrem estudos qualitativos (aspecto não considerado na versão original do instrumento).

Na etapa de validação, dois juízes avaliaram um conjunto de 25 RS e obtiveram um nível de concordância que variou de moderada à perfeita. Nota-se que o instrumento comportou-se de forma mais homogênea nesse segundo momento, possivelmente devido às modificações realizadas nos itens com menor grau de concordância no estudo piloto. A partir dos processos de adaptação e de validação realizados, a escala em sua versão final foi composta por 14 itens (Quadro 2) que conjuntamente indicam a pontuação geral do instrumento. Todavia, um dos pontos interessantes de se considerar é a possibilidade da avaliação item a item, visto que os critérios para cada um deles são independentes entre si<sup>11</sup>. Dessa forma, os itens tornam-se também ferramentas que guiam os pesquisadores no desenvolvimento de seus estudos de RS, atuando como um *checklist* de aspectos importantes a serem considerados. Cabe ponderar que não há, nesse sentido, a indicação de pontos de corte para a definição do que é uma boa ou uma má RS em termos metodológicos. Desde a sua gênese, a proposta do instrumento foi apontar critérios mínimos de avaliação, indicando que todos são necessários para revisões de boa qualidade. No entanto, entende-se que alguns itens podem ser mais importantes do que outros em determinadas situações<sup>11</sup>. Por exemplo, na Psicologia a discussão sobre o viés de publicação ainda é incipiente, mesmo que em outras áreas de conhecimento esse tema já esteja estabelecido<sup>20</sup>.

**Tabela 2.** Concordância interjuízes do estudo de validade.

Itens	Kappa	IC 95%	Concordância
1	0,62	[0,23, 1,00]	Substancial
2	0,61	[0,24, 0,97]	Substancial
3	0,46	[0,10, 0,83]	Moderada
4	0,42	[0,10, 0,74]	Moderada
5	0,66	[0,29, 1,0]	Substancial
6	0,46	[0,07, 0,85]	Moderada
7	0,42	[0,10, 0,74]	Moderada
8	0,50	[0,12, 0,89]	Moderada
9	0,52	[0,13, 0,90]	Moderada
10	0,44	[0,06, 0,82]	Moderada
11	0,47	[0,14, 0,80]	Moderada
12	0,43	[0,04, 0,82]	Moderada
13	0,86	[0,48, 1,00]	Quase perfeita
14	1,00	-	Perfeita

Ao se olhar para o campo da saúde coletiva, observa-se uma heterogeneidade de objetos e metodologias de pesquisa. Assim, é um desafio para o pesquisador escrever uma RS que considere as complexidades envolvidas nessa área. Para que isto ocorra de forma mais eficaz, é necessário que as pesquisas baseadas em evidências passem a adotar uma linha mais uniforme de apresentação e de publicação de seus resultados. Uma boa RS depende também do quão bons são os estudos revisados. Algumas vezes, um bom estudo deixa de fazer parte de uma RS, pois seus autores não utilizaram, por exemplo, palavras-chaves indexadas, dificultando que seu artigo seja acessado. Aliado a isso, a elaboração de um resumo claro, objetivo e detentor das principais informações do estudo torna-se fundamental para a seleção e a extração de dados a serem realizadas em uma RS.

Dessa forma, uma RS de qualidade, que traga dados sólidos sobre determinada temática e que embase decisões não apenas acadêmicas, mas também políticas, econômicas, educacionais e de saúde, depende da qualidade de todo o processo científico de pesquisa, escrita e publicação dos estudos. A metodologia apresentada neste trabalho pode constituir-se como um guia para pesquisadores, editores e pareceristas que buscam publicar revisões críticas de seus campos de atuação.

Apesar do cuidado com os processos de adaptação e de validação do instrumento, algumas limitações merecem ser apontadas. A quantidade de artigos avaliada pelos juízes pode ser considerada baixa, restringindo a variabilidade dos estudos incluídos e as análises realizadas. Além disso, não foi possível compará-lo com outro instrumento semelhante, pois não foram encontrados instrumentos no contexto brasileiro que avaliassem especificamente a qualidade metodológica das RS. Foram localizadas apenas escalas referentes à avaliação de estrutura e escrita formal de RS<sup>21</sup>. Cabe também pontuar que o funcionamento da escala ainda não foi avaliado após as modificações realizadas nos quatro itens indicados anteriormente. Assim, não há consenso de que eles apresentarão melhor comportamento do que suas versões anteriores.

Por fim, sugere-se que novos estudos sejam realizados para se buscar mais evidências de validade do instrumento. A indicação da pertinência dos itens e, em média, um grau satisfatório de concordância entre juízes sugerem que a escala é uma ferramenta útil e de fácil assimilação para o público acadêmico. Sua utilização pode promover ganhos metodológicos importantes nos estudos de RS realizados no Brasil, colaborando tanto com editores quanto com autores, a fim de se desenvolver trabalhos mais completos e robustos.

### Colaboradores

AB Costa, AP Zoltowski, SH Koller e MAP Teixeira participaram igualmente de todas as etapas de elaboração do artigo

## Referências

1. Glass G. Primary, secondary and meta-analysis of research. *Edu Res* 1976; 5(10):3-8.
2. Sousa M, Ribeiro A. Systematic review and meta-analysis of diagnostic and prognostic studies: A tutorial. *Arq Bras Cardiol* 2009; 92(3):241-251.
3. Cordeiro A, Oliveira GM, Rentería JM, Guimarães CA, Grupo de Estudo de RS do Rio de Janeiro. Revisão sistemática: uma revisão narrativa. *Rev Col Bras Cir* 2007; 34(6):428-431.
4. Francis G. Publication bias and the failure of replication in experimental psychology. *Psychon Bull Rev* 2012; 19(6):975-991.
5. Fernández-Ríos L, Buela-Casal G. Standards for the preparation and writing of Psychology review articles. *Int J Clin Health Psychol* 2009; 9(2):329-344.
6. Lopes A, Fracolli L. Revisão sistemática de literatura e metassíntese qualitativa: Considerações sobre sua aplicação na pesquisa em enfermagem. *Texto contexto - enferm* 2008; 17(1):771-778.
7. Papageorgiou S, Papadopoulos M, Athanasiou A. Evaluation of methodology and quality characteristics of systematic reviews in orthodontics. *Orthod Craniofac Res* 2011; 14(3):116-137.
8. Faggion CM, Listl S, Giannakopoulos NN. The methodological quality of systematic reviews of animal studies in dentistry. *Vet J* 2012; 193(2):140-147.
9. Mac.Donald SL, Canfield SE, Fesperman SF, Dahm P. Assessment of methodological quality of systematic reviews published in the urological literature from 1998 to 2008. *J Urol* 2010; 184(2):648-653.
10. Zoltowski APC, Costa AB, Teixeira MAP, Koller SH. Qualidade metodológica das revisões sistemáticas em periódicos de psicologia brasileiros. *Psic: Teor e Pesq* 2014; 30(1):107-114.
11. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009; 62(10):1013-1020.
12. Littell JH, Corcoran J, Pillai V. *Systematic reviews and meta-analysis*. New York: Oxford University Press; 2008.
13. Pearson A. Balancing the evidence: incorporating the synthesis of qualitative data into systematic reviews. *JBIM Reports* 2004; 2(2):45-64.
14. Jagannath V, Mathew JL, Asokan GV, Fedorowicz Z. Quality assessment of systematic reviews of health care interventions using AMSTAR. *Indian Pediatr* 2011; 48(5):383-385.
15. Shea BJ, Grimshaw JM, Wells GA, Boers M, AndeRson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7(10):1-7.
16. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; 44(11):1271-1278.
17. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316(8):450-455.
18. Urbina S. *Fundamentos da Testagem Psicológica*. Porto Alegre: Artmed; 2007.
19. Borsa JC, Damásio BF, Bandeira DR. Adaptação e validação de instrumentos psicológicos entre culturas: algumas considerações. *Paidéia* 2012; 22(53):423-432.
20. Yong E. Replication studies: bad copy. *Nature* 2012; 485(7398):298-300.
21. Padula RS, Pires RS, Alouche SR, Chiavegato LD, Lopes AD, Costa LOP. Análise da apresentação textual de Revisões sistemáticas em fisioterapia publicadas no idioma português. *Rev bras fisioter* 2012; 16(4):281-288.

Artigo apresentado em 18/04/2014

Aprovado em 04/09/2014

Versão final apresentada em 06/09/2014