

Adriana Lara Mesquita

Lanna Nicolau Teixeira

Thaís Gonçalves Lara

**INSTRUMENTOS DE MEDIDAS PARA CRIANÇAS E ADOLESCENTES COM  
SÍNDROME DE DOWN: uma revisão sistemática**

Belo Horizonte

Escola de Educação Física, Fisioterapia e Terapia Ocupacional/UFMG

2018

Adriana Lara Mesquita

Lanna Nicolau Teixeira

Thaís Gonçalves Lara

**INSTRUMENTOS DE MEDIDAS PARA CRIANÇAS E ADOLESCENTES COM  
SÍNDROME DE DOWN: uma revisão sistemática**

Monografia apresentada ao Curso de Graduação em Fisioterapia da Escola de Educação Física, Fisioterapia e Terapia Ocupacional da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Bacharel em Fisioterapia.

Orientadora: Profa. Ana Paula Bensemann Gontijo  
Co-Orientador: Ricardo Rodrigues de Sousa Junior

Belo Horizonte

Escola de Educação Física, Fisioterapia e Terapia Ocupacional/UFMG

2018

## RESUMO

Esta revisão tem como objetivo identificar os instrumentos de medidas disponíveis na literatura para a avaliação de crianças e adolescentes com Síndrome de Down durante o processo de reabilitação fisioterápica e analisar o atual nível de evidência das propriedades psicométricas desses instrumentos. Para isso, dois avaliadores independentes, estudaram nas bases de dados Medline, Lilacs e Scielo. Sete estudos foram selecionados pelos avaliadores. Tais estudos foram avaliados quanto a sua qualidade metodológica, através dos critérios COSMIN (Padrões Baseados no Consenso para a Seleção de Instrumentos de Medidas de Saúde) e quanto à qualidade das propriedades de medida e classificação da evidência, através dos critérios de Terwee. Os estudos investigaram seis instrumentos de medidas em crianças e adolescentes com Síndrome de Down. Destes instrumentos, quatro avaliaram funções motoras grossas (Basic Motor Skills, Gross Motor Function Measure, Timed Up and Go e Timed Up and Down Stairs) e dois equilíbrio (Four Square Step Test e Timed Standing Balance Test). A qualidade da evidência dos instrumentos variou de excelente a desconhecida, dependendo da propriedade de medida estudada. A escolha e administração destes instrumentos em crianças com SD devem ser feitas com cautela.

**Palavras-chave:** Instrumento de Medida. Síndrome de Down. Propriedades Psicométricas. Revisão Sistemática.

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>4</b>
<b>2. METODOLOGIA.....</b>	<b>5</b>
<b>3. RESULTADOS.....</b>	<b>7</b>
Descrição dos Instrumentos.....	11
Qualidade Metodológica dos Estudos.....	12
Qualidade das Propriedades Psicométricas e Síntese da Evidência.....	13
<b>4. DISCUSSÃO.....</b>	<b>17</b>
<b>5. CONCLUSÃO.....</b>	<b>20</b>
 <b>REFERÊNCIAS.....</b>	 <b>21</b>
 <b>ANEXOS.....</b>	 <b>26</b>

## 1 INTRODUÇÃO

Instrumentos padronizados de medida são essenciais tanto para a adequada avaliação do desenvolvimento de crianças com condições crônicas como a síndrome de Down (SD) como também para o planejamento e acompanhamento dos resultados das intervenções durante o tratamento fisioterápico (ELVRUM *et al.* 2016). Dentro do processo de reabilitação o uso de um adequado instrumento de medida é de suma importância para orientar profissionais da saúde, a família e o próprio paciente quanto ao seu prognóstico, limitações e potencialidades, contribuindo para a tomada decisão clínica (WRIGHT *et al.* 2013; MAJNEMER *et al.* 2002; LAW *et al.* 2002). Os resultados obtidos por instrumentos padronizados de medidas podem auxiliar também com decisões políticas sobre disposições de serviços e alocação de recursos dentro do sistema de saúde (WRIGHT *et al.* 2013).

Segundo Wright *et al.* (2013) diversos aspectos devem subsidiar a seleção de um instrumento de medida durante o processo de reabilitação fisioterapêutica. Um dos principais atributos a ser considerado é que o instrumento mensure constructos que são relevantes para o paciente e sua família. Para tal, é necessário que o instrumento aborde um ou mais domínios da Classificação Internacional de Funcionalidade (CIF) da Organização Mundial de Saúde (OMS) (FARIAS *et al.* 2005). A CIF tem sido uma ferramenta que tem ajudado os profissionais da reabilitação a refletir sobre as opções de intervenção nos diferentes domínios (estrutura e função de corpo, atividade e participação) considerando fatores pessoais e contextos ambientais. (NOVAK *et al.* 2013; SAMPAIO *et al.* 2010).

Uma segunda característica a ser considerada na seleção de um instrumento de medida é a qualidade de suas propriedades psicométricas, i.e, validade de conteúdo, confiabilidade teste-reteste e inter-examinadores e responsividade (WRIGHT *et al.* 2013; TERWEE *et al.* 2007). Além disso, o instrumento deve ser apropriado para cada população de interesse, considerando suas particularidades e diferentes condições de saúde. (MAJNEMER *et al.* 2002; WRIGHT *et al.*, 2013).

A SD é uma condição genética causada por alterações cromossômicas envolvendo o par do cromossomo 21, resultando em alterações físicas e mentais. (MENEGETTI *et al.* 2009). Considerando-se os domínios da CIF, dentre as disfunções observadas a nível de estrutura e função de corpo podemos citar a hipotonia muscular, frouxidão ligamentar, fraqueza muscular, déficits de equilíbrio e cognitivos associadas ou não à cardiopatia

congênita, disfunções neurológicas, endócrinas, auditivas e visuais, (MOREIRA *et al.* 2000; BORSSATTI *et al.* 2013). Em relação ao domínio atividade, indivíduos com SD apresentam atrasos na aquisição de habilidades motoras grossas e finas essenciais para um adequado desempenho das atividades de vida diária e, conseqüentemente para a adequada participação nas atividades escolares e sociais (TRINDADE *et al.* 2016).

Diversos instrumentos de medida de desfechos válidos e confiáveis têm sido utilizados por profissionais da área de reabilitação infantil, para avaliação e planejamento do tratamento de indivíduos com SD. No entanto, não foi encontrada na literatura uma revisão sistemática que resumisse os instrumentos utilizados. Portanto, o objetivo geral desta revisão sistemática foi identificar os instrumentos de medidas disponíveis para a avaliação de crianças e adolescentes com SD durante o processo de reabilitação e analisar o atual nível de evidência das propriedades psicométricas desses instrumentos.

## **2 METODOLOGIA**

O presente estudo seguiu o protocolo para realização de revisões de instrumentos de medida do COSMIN (*Consensus-based Standards for the selection of health Measurement Instruments*) (MOKKINK *et al.* 2010).

### 2.1 Estratégias de Busca e Seleção

As buscas pelos potenciais estudos foram realizadas nas bases de dados Medline, Lilacs e Scielo, no período de Dezembro de 2017 à Fevereiro de 2018. Para isso, foram utilizadas combinações booleanas, através das palavras chave “Down Syndrome”, “rehabilitation”, “assessment”, “outcomemeasure”, “psychometric”, “validity” e “reliability”. A busca e seleção foram feitas de forma independente por dois avaliadores utilizando um terceiro, em caso de divergência durante o consenso. As listas de referências foram analisadas para busca de outros possíveis estudos.

### 2.2 Critérios de Elegibilidade

Foram incluídos estudos metodológicos que avaliaram as propriedades de medida de instrumentos relacionados à avaliação motora de crianças e adolescentes com SD. Foram excluídos aqueles estudos que não avaliaram especificamente indivíduos com SD. Não houve restrição de data ou idioma.

### 2.3 Extração e Síntese dos Dados

A extração e síntese dos dados foram realizadas por dois examinadores de forma independente, utilizando um terceiro em caso de divergência, a fim de registrar as seguintes informações: características das amostras, objetivos dos estudos, propriedades de medida investigadas e principais resultados.

### 2.4 Avaliação da Qualidade Metodológica dos Estudos

Para avaliação qualidade metodológica dos estudos dois avaliadores independentes utilizaram o checklist COSMIN com a escala de 4 pontos (pobre, razoável, bom e excelente). Este checklist contém 9 domínios independentes (com 5 a 18 itens) para avaliação de estudos metodológicos que investigaram: consistência interna, confiabilidade, mensuração de erro, validade de conteúdo, validade estrutural, teste de hipóteses, adaptações culturais, validade de critério e responsividade. O escore final da qualidade metodológica é dado para cada domínio investigado no estudo considerando a pior pontuação (TERWEE *et al.* 2011). Para a avaliação da qualidade metodológica, foram seguidas adaptações realizadas por outras revisões de estudos metodológicos na área de reabilitação infantil onde foram desconsiderados os itens do checklist relacionados com o tamanho da amostra e itens perdidos (AMMANN-REIFFER *et al.* 2014; ELVRUM *et al.* 2016).

### 2.5 Avaliação da Qualidade das Propriedades Psicométricas

A qualidade das propriedades de medida de cada instrumento selecionado foi avaliada de acordo com os critérios propostos por Terwee *et al.* (2007). Cada propriedade de medida do instrumento é classificada em positiva (+), negativa (-) ou indeterminada (?) levando em consideração os resultados dos índices estatísticos (TERWEE *et al.* 2007). A tabela detalhada dos critérios para avaliação da qualidade das propriedades psicométricas encontra-se no anexo 1A.

### 2.6 Classificação do Nível da Evidência

O nível da evidência foi classificado em forte, moderado, limitado, conflitioso e desconhecido seguindo os critérios do Cochrane Back Review Group (VAN TULDER *et al.* 2003). A qualidade da evidência foi classificada levando em consideração a qualidade metodológica dos estudos, a qualidade das propriedades de medida e o número de estudos, conforme anexo 1B (ELVRUM *et al.* 2016; SCHELLINGERHOUT *et al.* 2012; VAN TULDER *et al.* 2003). O nível de evidência foi classificado para cada propriedade psicométrica de cada um dos instrumentos dos estudos selecionados por dois avaliadores independentes.

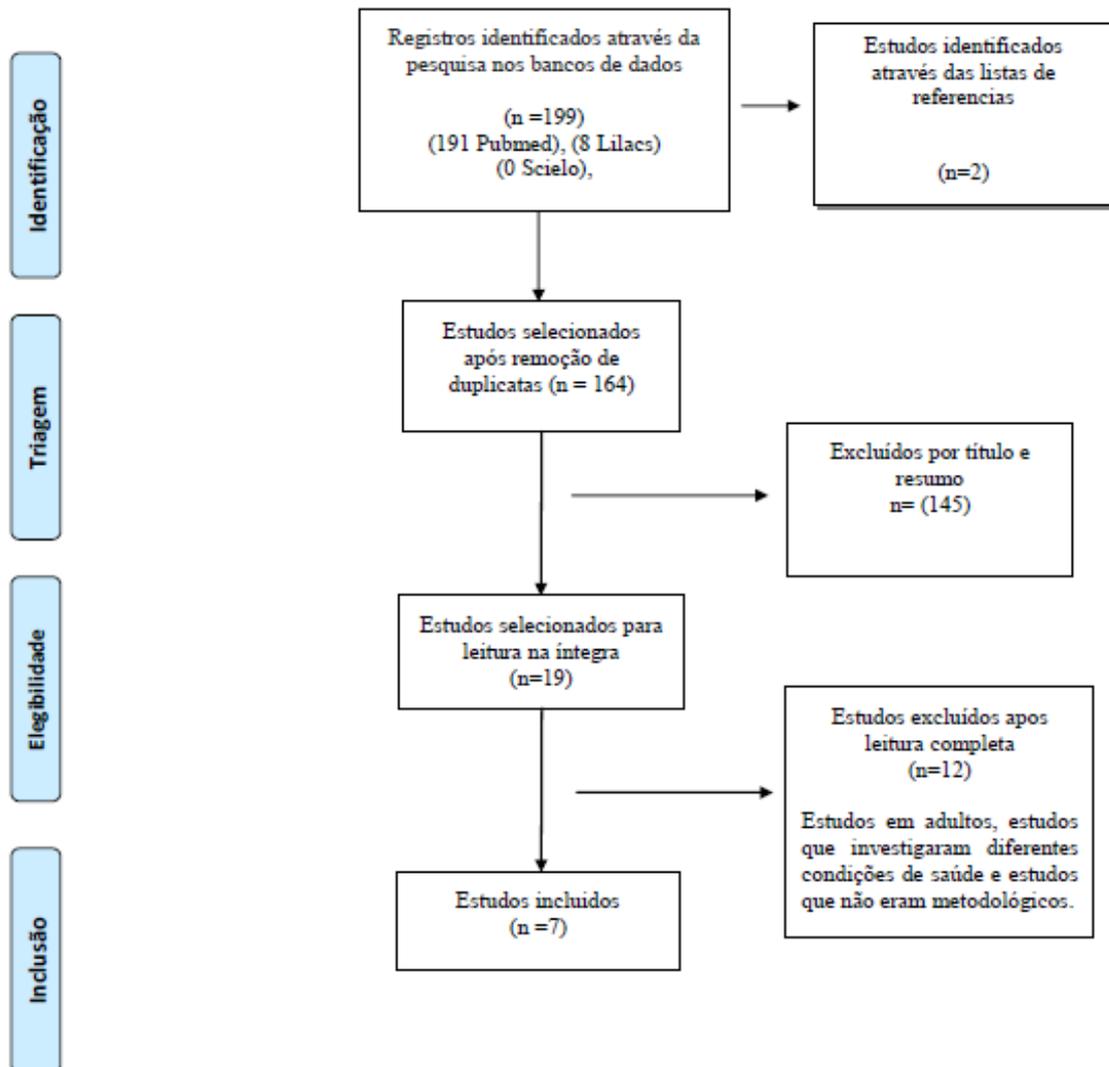
### **3 RESULTADOS**

#### 3.1 Seleção dos Estudos

A estratégia de busca identificou inicialmente 199 estudos, sendo 7 incluídos nesta revisão sistemática. O processo de busca e seleção dos estudos encontra-se detalhado na figura 1 no Diagrama de Seleção de Registros para Revisão Sistemática (PRISMA) (MOHER et al. 2009). O sumário das principais informações dos estudos selecionados encontra-se na tabela 1.

A média de idade dos participantes dos estudos incluídos foi 8,49 ( $\pm 4,61$ ). Foram totalizados 252 indivíduos, reportados 120 do sexo feminino e 120 do sexo masculino, com exceção apenas de um estudo que não informou o sexo dos participantes (MARTIN et al 2017). Os estudos selecionados utilizaram como critério de exclusão, indivíduos que possuíssem déficit visual não corrigido, comprometimento cognitivo grave e que sofreram algum tipo de lesão músculo-esquelética nos 6 meses anteriores ao início dos estudos.

Figura 1. Diagrama de Seleção de Registros para Revisão Sistemática



**Tabela 1. Sumário das Principais Informações dos Estudos Seleccionados**

Autor e Ano	Instrumento Estudado	População (Tamanho da amostra, média de idade e sexo)	Objetivo do Instrumento	Propriedades de Medida Investigadas			
				Validade de Critério	Confiabilidade	Mensurações de Erro	Responsividade
Van Del Heuvel <i>et al.</i> 2009	Basic Motor Skills (BMS)	n: 41 Idade: 2.02 anos ( $\pm 0.96$ ) 19 M 22 F	Avaliar habilidades motoras				X
Bandong <i>et al.</i> 2015	Four Square Step Test (FSST)	n: 14 Idade: 8.7 anos ( $\pm 1.9$ ) 5 M 9 F	Avaliar o equilíbrio dinâmico	X	X	X	
Verma <i>et al.</i> 2014		n: 13 Idade: 12.76 anos ( $\pm 2.9$ ) 8 M 5F		X	X		
Russell <i>et al.</i> 1998	Gross Motor Function Measure (GMFM)	n: 123 Idade: 2,39 anos 66 M 57 F	Avaliar função motora grossa		X		X
Aranha <i>et al.</i> 2016	Timed Standing Balance Test (TSBT)	n: 9 Idade: 13.5 anos ( $\pm 0.70$ ) 6 M 3F	Avaliar equilíbrio		X	X	
Martin <i>et al.</i>	Timed Up and	n: 12	Avaliar a				

2017	Down Stairs (TUDS)	Idade: 9.5 ( $\pm 4.4$ ) Sexo não informado	atividade de subir e descer escadas		X	X	
Martin <i>et al.</i> 2017	Timed Up and Go (TUG)	n: 12 Idade: 9.5 ( $\pm 4.4$ ) Sexo não informado	Avaliar mobilidade		X	X	
Nicolini <i>et al.</i> 2013		n: 40 Idade: 10.6 ( $\pm 4.4$ ) 16 M 24 F		X	X		

### 3.2 Descrição dos Instrumentos

Nos estudos selecionados foram analisadas as propriedades psicométricas de 6 instrumentos de medida. No domínio Estrutura e Função do Corpo da CIF, os estudos selecionados analisaram dois testes de equilíbrio: Timed Standing Balance Test (TSBT) e Four Square Test (FSST). No domínio Atividade foram analisados 4 instrumentos de medidas de funções motoras grossas: Basic Motor Skills (BMS), Gross Motor Function Measure (GMFM), Timed Up and Go (TUG) e Timed Up and Down Stairs (TUDS). Nenhum instrumento avaliou o domínio de participação da CIF. Abaixo a descrição dos testes localizados:

#### 3.2.1 Instrumentos no domínio “Estrutura e Função do Corpo”

O FSST é um teste de equilíbrio dinâmico que considera, como medida de desfecho, o tempo que o indivíduo realiza a tarefa de transpor obstáculos e de deslocar em quatro diferentes direções (DITE *et al.* 2002). As crianças realizam três tentativas, e a pontuação final é dada pela média (VERMA *et al.* 2014).

O TSBT é um teste de equilíbrio estático (indivíduo em um só pé no chão) e dinâmico (indivíduo em um só pé sobre uma almofada de espuma). São realizados quatro protocolos: estático com olhos abertos e olhos fechados, dinâmico com os olhos abertos e olhos fechados. Para cada um dos 4 protocolos, são realizadas 3 tentativas com cada perna, com intervalos de 30s e ordem de aplicação aleatória (EMERY *et al.* 2005).

#### 3.2.2 Instrumentos no domínio “Atividade”

O BMS é um instrumento de medida desenvolvido para monitorar mudanças nas funções motoras básicas de crianças com SD de 0 a 3 anos de idade constituído por 15 itens (LAUTESLAGER *et al.* 2004). É um teste referenciado em critério, ou seja, as crianças são comparadas com elas mesmas e não com um grupo controle (SCHOEMAKER *et al.* 2000).

O GMFM-88 é constituído de 88 itens divididos em cinco dimensões: (A) Deitar e Rolar, (B) Sentar, (C) Engatinhar e Ajoelhar, (D) Em Pé e (E) Andar, Correr, Pular. Cada item é pontuado em uma escala de zero a 3, gerando um escore para cada dimensão e um

escore total. Escores mais altos informam sobre maior repertório de função motora grossa da criança (RUSSELL *et al.* 2002; SCHIARITI *et al.* 2013).

O TUDS avalia a quantidade de tempo que os indivíduos despendem em realizar a tarefa de subir e descer 15 degraus com 20 cm de altura, usando o corrimão. Podem subir e descer os degraus alternando ou não os pés, usando ou não o corrimão ou até mesmo, saltitando (ZAINO *et al.* 2004).

O TUG é um teste no qual a criança sentada em uma cadeira, sem apoio das costas e braços tem que passar para de pé, caminhar por 3 metros e retornar à posição inicial. O tempo de realização do teste é cronometrado e está relacionado à mobilidade funcional (PODSIADLO *et al.*, 1991).

### 3.3 Qualidade Metodológica dos Estudos

A qualidade metodológica foi pontuada de acordo com a escala de 4 pontos do COSMIN nas seguintes propriedades de medida: validade de critério, confiabilidade, mensuração de erro (erro padrão de medida e mudança mínima detectável) e responsividade. A validade de critério tem o objetivo de investigar o quanto um instrumento se relaciona com outro já existente e tido como válido. (PORTNEY *et al.* 2009). A confiabilidade indica o quanto o instrumento é consistente em administrações repetidas (confiabilidade teste-reteste) e entre diferentes avaliadores (inter-examinadores). O erro padrão de medida indica a quantidade de erro aleatório presente na pontuação de um instrumento e o índice de mudança mínima detectável (MMD) indica sobre o mínimo de mudança que pode ocorrer além deste erro de medida. A interpretação destes índices é feita utilizando os índices de responsividade (PORTNEY *et al.*, 2009; TERWEE *et al.*, 2007). A responsividade de um instrumento indica sobre a capacidade deste em captar mudanças ao longo do tempo (PORTNEY *et al.* 2009). Os resultados da avaliação da qualidade metodológica dos estudos encontram-se na tabela 2.

A validade concorrente foi investigada em 3 estudos e teve sua qualidade metodológica pontuada entre “Pobre” a “Excelente” (BANDONG *et al.* 2015; NICOLINI *et al.* 2013; VERMA *et al.* 2014). A qualidade metodológica dos 6 estudos que analisaram os ICC teste-reteste e inter-examinadores variou entre “Pobre” e “Bom” (ARANHA *et al.* 2016; BANDONG *et al.* 2015; MARTIN *et al.* 2007; NICOLINI *et al.* 2013; RUSSEL *et al.* 1998; VERMA *et al.* 2014). Três estudos investigaram mensurações de erro dos instrumentos de medida estimando os valores do erro padrão de medida e índice de mudança mínima

detectável (ARANHA *et al.* 2016; BANDONG *et al.* 2015; MARTIN *et al.* 2007). A qualidade metodológica dos estudos variou de “Pobre” a “Bom”. A responsividade dos instrumentos de medida foi investigada por RUSSEL *et al.* (1998) e VAN DEL HEUVEL *et al.* (2009), e apresentaram qualidade metodológica “Pobre” e “Excelente”, respectivamente.

### 3.4 Qualidade das Propriedades Psicométricas e Síntese da Evidência

A qualidade das propriedades psicométricas dos instrumentos, bem como a classificação do nível da evidência encontram-se na tabela 3. A classificação do nível de evidência variou de forte à desconhecida devido ao pequeno número de estudos localizados, a qualidade metodológica dos estudos selecionados e a qualidade das propriedades psicométricas investigadas descritas nos tópicos abaixo:

#### 3.4.1 FSST

O teste FSST apresentou evidência moderada sobre a sua confiabilidade e limitada quanto à validade de critério e mensurações de erro devido à qualidade metodológica dos estudos (pobre a razoável) e a qualidade das propriedades de medida. As propriedades psicométricas do FSST foram analisadas em dois estudos em crianças de 6 a 17 anos (BANDONG *et al.* 2015; VERMA *et al.* 2014).

Em relação aos índices confiabilidade teste-reteste o FSST apresentou Coeficiente de Correlação Intra-Classe (CCI) de 0.70 e 0.81 (VERMA *et al.*, 2014; BANDONG *et al.*, 2015) e inter-examinadores de 0.78 (VERMA *et al.*, 2014). Ambos receberam avaliação positiva (+) para estas propriedades.

A validade concorrente foi investigada utilizando-se os testes Functional Reach Test (FRT) (VERMA *et al.*, 2014) e o TUG (BANDONG *et al.*, 2015) e a correlação encontrada foi moderada ( $\rho = -0.58$ ;  $P = 0.03$ ) e alta ( $\rho = 0.71$ ,  $p < 0.01$ ), respectivamente. Considerando-se o FRT a qualidade desta propriedade foi pontuada como indeterminada (?) uma vez que o teste FRT não é um teste padrão-ouro para avaliação do equilíbrio em crianças e adolescentes com SD. Por outro lado, a alta correlação com o TUG suporta o uso do FSST como uma medida de equilíbrio dinâmico, recebendo pontuação positiva (+).

Os valores de erro padrão de medida do instrumento, investigados por Bandong *et al.* (2015), foram de 2.32s e o índice MMD foi de 6.43s. Visto que não foram apresentados

outros índices de responsividade para a interpretação destes valores, a qualidade destas propriedades foi considerada indeterminada (?).

<b>Tabela 2, Qualidade Metodológica dos Estudos</b>					
<b>Autor e Ano</b>	<b>Instrumentos</b>	<b>Qualidade Metodológica</b>			
		<b>Validade de Critério</b>	<b>Confiabilidade</b>	<b>Mensurações de Erro</b>	<b>Responsividade</b>
Aranha <i>et al.</i> 2016	Timed Standing Balance Test (TSBT)		Razoável	Razoável	
Bandong <i>et al.</i> 2015	Four Square Step Test (FSST)	Razoável	Pobre	Razoável	
Martin <i>et al.</i> 2017	Timed Up and Go (TUG), Timed Up and Down Stairs (TUDS)		Bom	Bom	
Nicolini <i>et al.</i> 2013	Timed Up and Go (TUG)	Excelente	Pobre	Bom	
Russell <i>et al.</i> 1998	Gross Motor Function Measure (GMFM)		Bom		Pobre
Van Del Heuvel <i>et al.</i> 2009	Basic Motor Skills (BMS)				Excelente
Verma <i>et al.</i> 2014	Four Square Step Test (FSST)	Pobre	Bom		

### 3.4.2 TSBT

O TSBT foi aplicado em crianças e adolescentes entre 8 e 17 anos com SD (ARANHA *et al.* 2016). Apresentou evidência limitada tanto para a confiabilidade (qualidade metodológica do estudo foi classificada como razoável) quanto para as mensurações de erro (devido qualidade das propriedades de medida).

Os CCI teste-reteste para equilíbrio estático e dinâmico de 0.91 e 0.93 respectivamente, foram pontuados como qualidade positiva (+). O instrumento apresentou valores de erro padrão de medida de 0.21 a 0.59s e valores de MMD 0,58s a 1,63s. Estas propriedades foram avaliadas como indeterminadas (?) por não apresentarem os índices estatísticos adequados para interpretação desta propriedade.

### 3.4.3 BMS

A responsividade do BMS foi investigada em crianças com SD entre 3 meses e 3 anos de idade por Van Del Heuvel *et al.*, (2009). Este estudo indicou forte evidência sobre a capacidade do BMS de captar mudanças na função motora grossa após um período de 16 semanas. Para avaliação da responsividade do instrumento foi utilizado o Índice de Resposta de Guyatt (GRI), medida padronizada que indica a proporção dos escores de diferença observados durante um período de tempo (PORTNEY *et al.* 2009). O estudo encontrou valores de GRI = 2.55, indicando qualidade da propriedade de medida positiva (+). Além disso, o estudo comparou as mudanças entre crianças menores de 2 anos e crianças entre 2 e 3 anos. Crianças com SD menores que 2 anos de idade apresentaram uma mudança maior no escore médio de BMS em comparação com crianças com DS entre 2 a 3 anos de idade. Outras propriedades psicométricas deste instrumento ainda não foram estudadas.

### 3.4.4 GMFM

O GMFM apresenta evidência moderada quanto a confiabilidade e desconhecida quanto a responsividade em crianças com SD de zero a 6 anos de idade (RUSSEL *et al.* 1998). Os CCI inter-examinador e teste-reteste do GMFM-88 foram de 0.95 e 0.96 respectivamente, recebendo, uma avaliação positiva (+). No entanto, ao verificar a responsividade deste instrumento utilizando como critério externo o Bayley

Scale of Infant Development II (BSID-II), identificou-se que o GMFM-88 foi capaz de captar mudanças tal como o BSID-II, contudo a qualidade desta propriedade de medida foi classificada como indeterminada (?) uma vez que não foram apresentados no artigo índices estatísticos adequados para avaliação da responsividade.

#### 3.4.5 TUDS

O teste TUDS apresenta evidência moderada quanto à confiabilidade e conflituosa quanto as mensurações de erro em crianças de 3 a 17 anos com SD (MARTIN *et al.* 2007). O CCI teste-reteste do TUDS foi de 0,97, recebendo uma avaliação positiva (+). O instrumento apresentou um índice de MMD de 12,52s contudo, visto que não foram apresentados outros índices de responsividade para a interpretação deste valor, a qualidade desta propriedade psicométrica é indeterminada (?).

#### 3.4.6 TUG

O teste TUG apresentou nível de evidência moderado quanto à confiabilidade e conflituosa quanto à validade de critério e mensurações de erro ao ser administrado em crianças e adolescentes com SD de 3 a 18 anos (NICOLINI *et al.* 2013; MARTIN *et al.* 2017).

O instrumento apresenta boa confiabilidade teste-reteste, CCI: 0.82 (NICOLINI *et al.*, 2013) e 0.92 (MARTIN *et al.* 2017) apresentando qualidade positiva (+). A validade de critério do TUG foi analisada usando-se o GMFM como instrumento externo. A correlação encontrada foi baixa ( $r=0,49$ ,  $p= 0,001$ ) e, portanto, pontuada como negativa (-) (NICOLINI *et al.*, 2013). O MMD do instrumento foi de 1,26 segundos (MARTIN *et al.*, 2017) recebendo avaliação indeterminada (?) por não apresentar os índices estatísticos adequados para a interpretação deste valor.

<b>Tabela 3. Qualidade das Propriedades Psicométricas e Classificação do Nível da Evidência</b>				
<b>Instrumentos</b>	<b>Qualidade da Propriedade Psicométrica e Nível da Evidência</b>			
	<b>Validade de Critério</b>	<b>Confiabilidade</b>	<b>Mensurações de Erro</b>	<b>Responsividade</b>
Basic Motor Skills (BMS)				Positiva (+)
				<b>Forte</b>
Four Square Step Test (FSST)	Indeterminada (?)	Positiva (+) (+)	Indeterminada (?)	
	<b>Limitada</b>	<b>Moderada</b>	<b>Limitada</b>	
Gross Motor Function Measure (GMFM)		Positiva (+)		Indeterminada (?)
		<b>Moderada</b>		<b>Desconhecida</b>
Timed Standing Balance Test (TSBT)		Positiva (+)	Indeterminada (?)	
		<b>Limitada</b>	<b>Limitada</b>	
Timed Up and Go (TUG),	Negativa (-)	Positiva (+) (+)	Indeterminada (?)	
	<b>Conflituosa</b>	<b>Moderada</b>	<b>Conflituosa</b>	
Timed Up and Down Stairs (TUDS)		Positiva (+)	Indeterminada (?)	
		<b>Moderada</b>	<b>Conflituosa</b>	

## 4 DISCUSSÃO

Esta é a primeira revisão sistemática que identifica os instrumentos de medida disponíveis para profissionais da área de reabilitação motora para avaliação de crianças e adolescentes com SD. Identificamos 6 instrumentos nos quais as propriedades psicométricas foram examinadas. Dos instrumentos encontrados, considerando-se o modelo proposto pela CIF, dois abordaram aspectos de “estrutura e função do corpo” e quatro avaliaram o desenvolvimento motor grosso, constructo do nível “atividade”. Nenhum dos instrumentos de medidas de desfechos avaliou constructos relacionados à participação do indivíduo.

No nível de estrutura e função do corpo da CIF os dois instrumentos utilizados em crianças com SD apresentam, como medida de desfecho, a mensuração do equilíbrio em atividades na postura de pé. Em relação às propriedades psicométricas, a validade concorrente foi investigada somente para o FSST, com evidências limitadas. Não foram identificadas diferenças entre os 2 instrumentos nas qualidades das propriedades psicométricas de confiabilidade e erro de medida. No entanto, considerando-se as medidas de confiabilidade o FSST apresentou melhor nível de evidência devido a qualidade metodológica dos estudos. Não foram encontrados na literatura estudos que avaliassem o controle postural em crianças com SD que ainda não adquiriram a postura ortostática. Visto que o equilíbrio é um componente comumente alterado nesta população (MOREIRA *et al.* 2000; BORSSATTI *et al.* 2013), esta revisão identificou a necessidade de desenvolvimento de pesquisas que investiguem instrumentos que contemplem componentes importantes na avaliação do equilíbrio em crianças menores, tais como reações de retificação e equilíbrio nas posturas supino, prono e sentado.

Um segundo aspecto evidenciado nesta revisão foi a ausência de estudos que examinassem as propriedades dos instrumentos para mensuração do alinhamento articular, força muscular, condicionamento cardiorrespiratório, variáveis que se encontram comumente deficitárias nesta população (MOREIRA *et al.* 2000; BORSSATTI *et al.* 2013). Estudo que investigou condicionamento cardiorrespiratório em adultos jovens com SD utilizando o teste de caminhada de 6 minutos como medida de desfecho, indicou que o instrumento apresenta índices adequados de confiabilidade e

que o teste foi capaz de captar mudanças no condicionamento cardiorrespiratório desta população (CASEY *et al.*, 2012).

Quatro instrumentos do domínio “atividade” da CIF foram localizados nesta revisão. O BMS e o GMFM avaliam funções motoras grossas e o TUDS e TUG, consideram o tempo em que as crianças realizam atividades dinâmicas. Dos quatro instrumentos, o BMS foi o único desenvolvido especificamente para crianças com SD. No entanto, a única propriedade avaliada foi a responsividade. Portanto, a sua indicação como instrumento de medida deve ser realizada com cautela uma vez que ele não possui outras propriedades psicométricas investigadas, como por exemplo, validade e confiabilidade. Um segundo instrumento encontrado na revisão para avaliação da função motora foi o GMFM. Embora desenvolvido para avaliação de crianças com paralisia cerebral, o GMFM teve suas medidas de confiabilidade e responsividade investigadas em crianças com SD (RUSSEL, 1998). No entanto, apenas duas medidas psicométricas foram examinadas, a confiabilidade e a responsividade. Os resultados indicaram que o teste apresenta bons índices de confiabilidade. Em relação a responsividade, a classificação da evidência para esta propriedade é desconhecida, uma vez que no estudo não foram apresentados os índices para cálculo desta propriedade como sugerido por Terwee *et al.* (2007).

Os instrumentos TUG e TUDS apresentaram resultados semelhantes em relação às propriedades de confiabilidade e medida de erro. No entanto, como informações sobre mudanças clinicamente importantes não foram reportadas nos estudos incluídos, o nível de evidência sobre a medida de erro dos instrumentos permanece indeterminado. Portanto, resultados sobre medida de erro destes instrumentos devem ser interpretados com cautela. Além da confiabilidade e medida de erro, o TUG teve sua validade de critério avaliada. Apesar da boa qualidade metodológica do estudo, o nível da evidência foi classificado como conflituooso uma vez que os resultados indicaram uma baixa correlação do TUG com o GMFM.

Os quatro instrumentos localizados no domínio “atividade” avaliam as funções motoras em um ambiente padronizado (capacidade). Embora instrumentos que mensuram a execução de atividades em um contexto real (desempenho) como o Inventário de Avaliação da Incapacidade Pediátrica (PEDI) e a Medida de Independência Funcional (WeeFIM) (MANCINI *et al.* 2005; MSALL *et al.* 1994) já terem sido utilizados para caracterizar o desempenho funcional em crianças com SD

(DOLVA *et al.* 2004; LIN *et al.* 2016; MANCINI *et al.* 2003) suas propriedades psicométricas não foram investigadas para indivíduos com SD.

Esta revisão indica que as propriedades de medida ainda precisam ser determinadas para a maioria dos instrumentos e diferentes diagnósticos. Não surpreendentemente, a confiabilidade foi a medida mais comumente avaliada. No entanto, para medidas avaliativas, informações sobre concordância, a qual é quantificada com a medida de erro, são muito mais robustas do que a confiabilidade. A medida de erro, por ser expressa na unidade da medida do instrumento, permite a interpretação do índice de mudança clinicamente importante (MCID). Este índice informa o mínimo de mudança detectada por um instrumento que é clinicamente significativa para o paciente (DE VET *et al.* 2006; LEOPOLD *et al.* 2017). A classificação do nível de evidência dos instrumentos de avaliação encontrados nesta revisão quanto a mensurações de erro foi limitada ou conflituosa, pois nenhum dos estudos selecionados forneceu índices estatísticos que permitissem a interpretação destes valores. Portanto, estudos futuros devem investigar índices de mudanças clinicamente importantes nos instrumentos de medida.

O nível da evidência do uso dos instrumentos localizados variou de desconhecido a forte. Visto que poucos estudos tiveram avaliação “excelente” quanto aos aspectos metodológicos, estudos com melhor qualidade poderão melhorar a classificação da evidência dos instrumentos localizados. Além disso, todos os instrumentos apresentaram alguma limitação quanto as suas propriedades psicométricas. Instrumentos com boas propriedades psicométricas são de suma importância para uma avaliação adequada, e, conseqüentemente, para o acompanhamento apropriado do indivíduo durante o processo de reabilitação (TERWEE *et al.* 2011; ELVRUM *et al.* 2016).

Uma possível limitação deste estudo foi a quantidade pequena de bases de dados pesquisadas, devido a indisponibilidade de acesso. Além disso, optou-se pela retirada do item “tamanho amostral” do instrumento COSMIN durante a avaliação da qualidade metodológica, devido à escassez de estudos na população pediátrica com grande tamanho amostral.

## **5 CONCLUSÃO**

Neste estudo foi conduzida uma revisão sistemática para identificação de instrumentos disponíveis na literatura para avaliação de crianças e adolescentes com SD. Seis instrumentos foram localizados nesta revisão, 2 avaliaram aspectos do domínio estrutura e função do corpo e 4 do domínio atividade da CIF. Não foram encontrados estudos que avaliassem o domínio participação. Dos estudos que avaliaram o domínio atividade, nenhum dos instrumentos utilizados avaliou o desempenho. A escolha e administração destes instrumentos em crianças com SD devem ser feitas com cautela, pois todos apresentaram limitações em suas propriedades psicométricas quando aplicados nesta população. Futuros estudos devem melhorar a qualidade da evidência destes instrumentos.

## REFERÊNCIAS

ALMEIDA KM, ALBUQUERQUE KA, FERREIRA, ML, AGUIAR SK, MANCINI MC. Reliability of the Brazilian Portuguese version of the Gross Motor Function Measure in children with cerebral palsy. **Brazilian Journal of Physical Therapy**, v. 20, n.1, p 73-80, 2016.

AMMANN-REIFFER, Corinne *et al.* Measurement properties of gait-related outcomes in youth with neuromuscular diagnoses: a systematic review. **Physicaltherapy**, v. 94, n. 8, p. 1067-1082, 2014.

ARANHA, VencitaPriyanka; SAMUEL, Asir John; SAXENA, Shikha. Reliability and sensitivity to change of the timed standing balance test in children with down syndrome. **Journal of neurosciences in rural practice**, v. 7, n. 1, p. 77, 2016.

BANDONG, AilaNica J.; MADRIAGA, Gilbert O.; GORGON, Edward James R. Reliability and validity of the Four Square Step Test in children with cerebral palsy and Down syndrome. **Research in developmental disabilities**, v. 47, p. 39-47, 2015.

BORSSATTI, Francieli; DOS ANJOS, Francine Batista; RIBAS, Danieli Isabel Romanovitch. Efeitos dos exercícios de força muscular na marcha de indivíduos portadores de Síndrome de Down. **Fisioterapia em Movimento**, v. 26, n. 2, 2013.

DE VET HC, TERWEE CB, KNOL DL, BOUTER LM. When to use agreement versus reliability measures. **Journal of Clinical Epidemiology** v. 59, p.1033–1039, 2006.

DITE, Wayne; TEMPLE, Vivienne A. A clinical test of stepping and change of direction to identify multiple falling older adults. **Archives of physical medicine and rehabilitation**, v. 83, n. 11, p. 1566-1571, 2002.

DOLVA, Anne-Stine; COSTER, Wendy; LILJA, Margareta. Functional performance in children with Down syndrome. **American journal of occupational therapy**, v. 58, n. 6, p. 621-629, 2004.

ELVRUM, Ann-Kristin G. *et al.* Outcome measures evaluating hand function in children with bilateral cerebral palsy: a systematic review. **Developmental Medicine & Child Neurology**, v. 58, n. 7, p. 662-671, 2016.

EMERY, Carolyn A. *et al.* Development of a clinical static and dynamic standing balance measurement tool appropriate for use in adolescents. **Physical therapy**, v. 85, n. 6, p. 502-514, 2005.

FARIAS, Norma; BUCHALLA, Cassia Maria. A classificação internacional de funcionalidade, incapacidade e saúde da organização mundial da saúde: conceitos, usos e perspectivas. **Revista brasileira de epidemiologia**, v. 8, p. 187-193, 2005.

GOH, Esther Y. *et al.* Reliability and concurrent validity of four square step test scores in subjects with chronic stroke: a pilot study. **Archives of physical medicine and rehabilitation**, v. 94, n. 7, p. 1306-1311, 2013.

LAUTESLAGER, P. E. M., PENNING, A. H., VERMEER, A., HELDERS, P. J. M., & Hart, H 't. Test van 'Basis-motor is chevaardigheden van kinderen met het syndroom van Down': onderzoeknaarbetroikbaarheid en constructvaliditeit. **NederlandsTijdschriftvoorFysiotherapie**, v.108, n.6, p.155-163, 1998.

LAUTESLAGER, Peter EM. Children with Down's syndrome. **Motor development and intervention**. Amesfoort, the Netherlands: sHeeren Loo Zorggroep, 2004.

LAW, Mary *et al.* Measuring outcomes in children's rehabilitation: A decision protocol. **Archives of physical medicine and rehabilitation**, v. 80, n. 6, p. 629-636, 1999.

LEOPOLD S, Porcher R. Editorial: The Minimum Clinically Important Difference- The Least We Can Do. **Clinical Orthopaedics and Related Research.**, v.475, n.4, p.929-932, 2017.

LIBERATTI, Alessandro *et al.* Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Published: July 21, 2009.

LIN, Hsiang-Yu *et al.* Functional independence of Taiwanese children with Down syndrome. **Developmental Medicine & Child Neurology**, v. 58, n. 5, p. 502-507, 2016.

MAJNEMER, Annette; LIMPEROPOULOS, Catherine. Importance of outcome determination in pediatric rehabilitation. **Developmental medicine and child neurology**, v. 44, n. 11, p. 773-777, 2002.

MANCINI, Marisa Cotta *et al.* Comparison of functional performance among children with Down Syndrome and children with age appropriate development at 2 and 5 years of age. **Arq Neuropsiquiatr** v.61, n.2-B, p.409-415, 2003.

MARTIN, Kathy *et al.* Development and reliability of an observational gait analysis tool for children with Down syndrome. **Pediatric Physical Therapy**, v. 21, n. 3, p. 261-268, 2009.

MARTIN, Kathy *et al.* Minimal detectable change for TUG and TUDS tests for children with down syndrome. **Pediatric Physical Therapy**, v.29, p.77-82, 2017.

MENEGHETTI, Cristiane Helita Zorel *et al.* Avaliação do equilíbrio estático de crianças e adolescentes com síndrome de Down. **Brazilian Journal of Physical Therapy/ Revista Brasileira de Fisioterapia**, v. 13, n. 3, 2009.

MOHER, David *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **PLoS medicine**, v. 6, n. 7, p. e1000097, 2009.

MOKKINK, L. *et al.* International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient reported outcomes: results of the COSMIN study. **J ClinEpidemiol.**, v.63, p.737-745, 2010.

MOREIRA, Lília Maria de Azevedo; EL-HANI, Charbel Niño; GUSMÃO, Fábio Alexandre Ferreira. **A síndrome de Down e sua patogênese**: considerações sobre o determinismo genético. 2000.

MSALL, Michael E. *et al.* The Functional Independence Measure for Children (WeeFIM) conceptual basis and pilot use in children with developmental disabilities. **Clinical Pediatrics**, v. 33, n. 7, p. 421-430, 1994.

NICOLINI□PANISSON, Renata D.'Agostini; DONADIO, Márcio VF. Normative values for the Timed 'Up and Go'test in children and adolescents and validation for individuals with Down syndrome. **Developmental Medicine & Child Neurology**, v. 56, n. 5, p. 490-497, 2014.

NOVAK, Iona *et al.* A systematic review of interventions for children with cerebral palsy: state of the evidence. **Developmental Medicine & ChildNeurology**, v. 55, n. 10, p. 885-910, 2013.

NUBILA, Di; VENTURA, Heloisa Brunow; BUCHALLA, Cassia Maria. O papel das Classificações da OMS-CID e CIF nas definições de deficiência e incapacidade. **Revista Brasileira de Epidemiologia**, v. 11, p. 324-335, 2008.

PORTNEY, Leslie Gross *et al.* **Foundations of clinical research: applications to practice**. 2009.

PODSIADLO, Diane; RICHARDSON, Sandra. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. **Journal of the American geriatrics Society**, v. 39, n. 2, p. 142-148, 1991.

RUSSELL, Dianne *et al.* Evaluating motor function in children with Down syndrome: validity of the GMFM. **Developmental Medicine & Child Neurology**, v. 40, n. 10, p. 693-701, 1998.

RUSSELL, D. J., ROSENBAUM, P. L., AVERY, L. M., & LANE, M. Gross motor function measure (GMFM-66 & GMFM-88). **User's manual**. Suffolk: Mac Keith Press, 2002.

RUSSELL, D. J., ROSENBAUM, P. L., AVERY, L. M., & LANE, M. (2002). **Medida da função motora grossa: (GMFM - 66 & GMFM - 88): manual do usuário**. Cyrillo ST, Galvão MCS, tradutores. São Paulo: Memnon, 2011.

SAMPAIO, Rosana F. *et al.* Aplicação da classificação internacional de funcionalidade, incapacidade e saúde (CIF) na prática clínica do fisioterapeuta. **Rev Bras Fisioter**, v. 9, n. 2, p. 129-36, 2005.

SCHELLINGERHOUT, Jasper M. *et al.* Measurement properties of translated versions of neck-specific questionnaires: a systematic review. **BMC Medical Research Methodology**, v. 11, n. 1, p. 87, 2011.

SCHIARITI V, Klassen AF, Cieza A, Sauve K, O'Donnell M, Armstrong R, Mâsse L . Comparing Contents of Outcome Measures in Cerebral Palsy Using the International Classification of Functioning (ICF-CY): A Systematic Review. **European Journal Paediatric Neurology Society**. 2013; 1-12

SCHOEMAKER, M., Ketelaar, M., & SMITS-ENGELSMAN, B. C. M. (2000). Meetinstrumenten voor de motorische ontwikkeling van kinderen. *In*: R. VAN EMPELEN, R. NIJHUIS-VAN DER SANDEN, & A. HARTMAN (red.), **Kinderfysiotherapie. Maarssen: Elsevier gezondheidszorg**. p. 147–158.

TERWEE CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. **J ClinEpidemiol** v.60, p.34-42, 2007. MARTINS, Gilberto A. *et al.* Sobre Confiabilidade e Validade. **RBGN**, São Paulo, v. 8, n. 20, p. 1-12, jan. / abr. 2006.

TERWEE CB, MOKKINK LB, KNOL DL, OSTELO RWJG, BOUTER LM, DE VET HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. **Quality of Life Research** July 6, 2011.

TRINDADE, André Soares; NASCIMENTO, Marcos Antonio do. Avaliação do Desenvolvimento Motor em Crianças com Síndrome de Down. **Rev. bras. educ. espec.**, Marília, v. 22, n. 4, p. 577-588, Dec. 2016 .

VAN DEN HEUVEL, Marieke E. *et al.* responsiveness of the test of basic motor skills of children with Down Syndrome. **Physical & occupational therapy in pediatrics**, v. 29, n. 1, p. 71-85, 2009.

VAN TULDER, Maurits *et al.* Updated method guidelines for systematic reviews in the cochrane collaboration back review group. **Spine**, v. 28, n. 12, p. 1290-1299, 2003.

VERMA, Ajai; SAMUEL, Asir John; ARANHA, VencitaPriyanka. The four square step test in children with Down syndrome: Reliability and concurrent validity. **Journal of pediatric neurosciences**, v. 9, n. 3, p. 221, 2014.

WRIGHT, Virginia *et al.* Motor measures: A moving target?.In: **Seminars in pediatric neurology**. Elsevier, 2013. p. 84-99.

ZAINO, Christopher A.; MARCHESE, Victoria Gocha; WESTCOTT, Sarah L. Timed up and down stairs test: preliminary reliability and validity of a new measure of functional mobility. **Pediatric Physical Therapy**, v. 16, n. 2, p. 90-98, 2004.

## ANEXO IA- Tabela de Avaliação da Qualidade das Propriedades de Medidas

Table 1  
Quality criteria for measurement properties of health status questionnaires

Property	Definition	Quality criteria <sup>a,b</sup>
1. Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	+A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection; ?A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method; -No target population involvement; 0No information found on target population involvement.
2. Internal consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	+Factor analyses performed on adequate sample size (7 * # items and $\geq 100$ ) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95; ?No factor analysis OR doubtful design or method; -Cronbach's alpha(s) $< 0.70$ or $> 0.95$ , despite adequate design and method; 0No information found on internal consistency.
3. Criterion validity	The extent to which scores on a particular questionnaire relate to a gold standard	+Convincing arguments that gold standard is "gold" AND correlation with gold standard $\geq 0.70$ ; ?No convincing arguments that gold standard is "gold" OR doubtful design or method; -Correlation with gold standard $< 0.70$ , despite adequate design and method; 0No information found on criterion validity.
4. Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	+Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses; ?Doubtful design or method (e.g., no hypotheses); -Less than 75% of hypotheses were confirmed, despite adequate design and methods; 0No information found on construct validity.
5. Reproducibility		
5.1. Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	+MIC $< SDC$ OR MIC outside the LOA OR convincing arguments that agreement is acceptable; ?Doubtful design or method OR (MIC not defined AND no convincing arguments that agreement is acceptable); -MIC $\geq SDC$ OR MIC equals or inside LOA, despite adequate design and method; 0No information found on agreement.
5.2. Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	+ICC or weighted Kappa $\geq 0.70$ ; ?Doubtful design or method (e.g., time interval not mentioned); -ICC or weighted Kappa $< 0.70$ , despite adequate design and method; 0No information found on reliability.
6. Responsiveness	The ability of a questionnaire to detect clinically important changes over time	+SDC or SDC $< MIC$ OR MIC outside the LOA OR RR $> 1.96$ OR AUC $\geq 0.70$ ; ?Doubtful design or method; -SDC or SDC $\geq MIC$ OR MIC equals or inside LOA OR RR $\leq 1.96$ OR AUC $< 0.70$ , despite adequate design and methods; 0No information found on responsiveness.
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	+ $\leq 15\%$ of the respondents achieved the highest or lowest possible scores; ?Doubtful design or method; - $> 15\%$ of the respondents achieved the highest or lowest possible scores, despite adequate design and methods; 0No information found on interpretation.
8. Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	+Mean and SD scores presented of at least four relevant subgroups of patients and MIC defined; ?Doubtful design or method OR less than four subgroups OR no MIC defined; 0No information found on interpretation.

MIC = minimal important change; SDC = smallest detectable change; LOA = limits of agreement; ICC = Intraclass correlation; SD, standard deviation.

<sup>a</sup> + = positive rating; ? = indeterminate rating; - = negative rating; 0 = no information available.

<sup>b</sup> Doubtful design or method = lacking of a clear description of the design or methods of the study, sample size smaller than 50 subjects (should be at least 50 in every (subgroup) analysis), or any important methodological weakness in the design or execution of the study.

Retirado de "Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res.* 2012; 21 p 663.

## Anexo IB- Tabela de Critérios de Classificação da Qualidade da Evidência

**Table 1** Levels of evidence for the overall quality of the measurement property [17]

Level	Rating	Criteria
Strong	+++ or ----	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or --	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or -	One study of fair methodological quality
Conflicting	±	Conflicting findings
Unknown	?	Only studies of poor methodological quality

[..] reference number, + positive result, - negative result

Retirado de “ Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res.* 2012; 21 p 662.