



UNIVERSIDADE
LUSÓFONA

Ana Catarina da Silva Oliveira

**Avaliação das Funções Executivas em
Crianças através de
Questionários/Inventários: Uma Revisão
Sistemática**

Trabalho realizado sob orientação da
Professora Doutora Marisa Filipe

Março 2023



UNIVERSIDADE
LUSÓFONA

Ana Catarina da Silva Oliveira

**Avaliação das Funções Executivas em
Crianças através de
Questionários/Inventários: Uma Revisão
Sistemática**

Tese de Mestrado em Psicologia Clínica e da Saúde

Tese defendida em provas públicas na Universidade
Lusófona do Porto no dia 23/03/2023 perante o júri
seguinte:

Presidente: Prof. Doutor Diogo Lamela

Arguente: Prof. Doutora Rita Antunes (Universidade
Lusíada)

Orientadora: Prof. Doutora Marisa Filipe

Março 2023

É autorizada a reprodução integral desta tese/dissertação apenas para efeitos de investigação, mediante declaração escrita do interessado, que a tal se compromete.

Agradecimentos

Exprimo, desta forma, os meus agradecimentos àqueles que contribuíram, de forma relevante, para a elaboração desta dissertação de mestrado.

Em primeiro lugar, gostaria de agradecer à Professora Doutora Marisa Filipe. Agradeço por toda a orientação, pelo apoio e acompanhamento prestados, fundamental para a conclusão deste trabalho.

Agradeço, de igual forma, à minha família e amigos por todo o apoio, força e carinho que sempre me prestaram ao longo de todo o meu percurso académico.

Mas também agradeço a todos que se cruzaram comigo ao longo desta trajetória, especialmente professores e colegas que, diretamente ou indiretamente, contribuíram para a realização final desta tese, pelas competências que me transmitiram e pela partilha de ideias e conhecimentos.

A todos, muito obrigada!

Resumo

A avaliação das funções executivas é essencial devido à sua importância para uma vida bem-sucedida e adaptada, principalmente para a primeira infância, fase em que estas começam a desenvolver-se (Anderson et al., 2001). Assim, a presente revisão pretende sistematizar os questionários e inventários existentes para a avaliação das funções executivas em crianças até aos doze anos de idade, com o intuito de contribuir para a aplicabilidade dos mesmos. Consequentemente, o estudo debruça-se sobre a análise das características psicométricas dos instrumentos e na qualidade dos seus estudos, através de uma análise de risco de viés.

De forma a garantir o rigor metodológico, esta revisão sistemática segue as diretrizes *PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)*; Page et al., 2021). A pesquisa de artigos foi conduzida através do motor de busca *EBSCO*, tendo como referência os últimos dez anos e recorrendo às seguintes bases de dados: *Psychology and Behavioral Sciences, Academic Search Complete, APA PsycInfo e Eric*. Após a fase de elegibilidade, seguiram para inclusão no estudo 83 artigos, por respeitarem os critérios definidos. Os resultados sugerem uma grande diversidade de questionários/inventários (cerca de 18) que avaliam as FE para diversas faixas etárias (dos 2-21 anos), sendo o *Behaviour Rating Inventory of Executive Function (BRIEF)* o mais frequentemente usado, seguido do *Behavior Assessment System for Children, Second Edition (BASC-2)* e do *Childhood Executive Functioning Inventory (CHEXI)*. Além disso, os resultados revelam boas características psicométricas dos instrumentos, mas revelam pontuações desfavoráveis no que toca à qualidade dos seus estudos.

Em conclusão, o estudo sobre este tema permite notabilizar a necessidade de maior atenção no que diz respeito à investigação nesta área.

Palavras-Chave: Avaliação; Funções Executivas; Questionários; Inventários; Crianças

Abstract

The assessment of executive functions is essential due to its importance for a well-accepted and adapted life, especially for early childhood, when they begin to develop (Anderson et al., 2001). Thus, the review intends to systematize the present dates and existing inventories until the assessment of executive functions in children, in order to contribute to their applicability. Consequently, the study focuses on the analysis of the psychometric characteristics of the instruments and the quality of their studies, through a risk of bias analysis.

In order to review the methodological rigor, this systematic follows the PRISMA guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Page et al., 2021). The article search was searched EBSCO and through the data engine, having as reference the last ten years and using the following databases: Psychology Behavioral Sciences, Academic Search Complete, APA PsycInfo and Eric. After the eligibility phase, they were included in study 83, as they met the defined criteria. The results suggest a great diversity of questionnaires/inventories (about 18) that assess the EFs for different age groups (from 2-21 years old), being the Behavior Rating Inventory of Executive Function (BRIEF) was the most frequently used, followed by the Behavior Assessment System for Children, Second Edition (BASC-2) and the Childhood Executive Functioning Inventory (CHEXI). In addition, the results reveal good psychometric characteristics of the instruments, but reveal unfavorable scores regarding the quality of their studies.

In conclusion, the study on this topic makes it possible to highlight the need for greater attention with regard to research in this area.

Keywords: Evaluation; Executive Functions; Questionnaires; Inventories; Children

Índice

1. Introdução Teórica.....	1
1.1. Componentes Executivas.....	2
1.2. Modelos Teóricos das FE.....	3
1.3. Desenvolvimento das FE.....	4
1.4. Alterações nas FE.....	6
1.5. Avaliação das FE.....	7
1.6. Objetivos de Investigação.....	9
2. Método.....	10
2.1. Pesquisa.....	10
2.1.1. Questão de Investigação.....	10
2.1.2. Base de Dados.....	10
2.1.3. Termos de Pesquisa.....	10
2.1.4. Critérios de Elegibilidade.....	10
2.2. Procedimentos.....	11
2.2.1. Recolha de Dados.....	11
2.3. Estratégia de Análise de Dados.....	11
3. Resultados.....	12
3.1. Características Gerais dos Estudos.....	13
3.2. Questionários/Inventários de FE.....	14
3.2.1. Características dos Instrumentos.....	16
3.3. Propriedades de Medida.....	20
4. Discussão.....	32
4.1. Discussão dos principais resultados.....	32
4.1.1. Características dos estudos/instrumentos.....	32
4.1.2. Propriedades de Medida.....	33
4.1.3. Ferramenta <i>Cosmin Checklist</i>	34
4.2. Limitações e Implicações Clínicas.....	34
5. Conclusão.....	36

Índice de Tabelas

Tabela 1. Visão Geral dos Instrumentos.....	15
Tabela 2. Características dos Instrumentos.....	17
Tabela 3. Risco de Viés.....	21
Tabela 4. Classificação do Risco de Viés.....	27

Índice de Figuras

Figura 1. Fluxograma PRISMA 2020.....	12
---------------------------------------	----

Índice de Anexos

Anexo 1. Taxonomia de propriedades de medida do *Cosmin*

Anexo 2. Método de Classificação das Propriedades de Medida

Anexo 3. Lista de Verificação *Cosmin Risk of Bias*

Lista de Abreviaturas

BADS-C	<i>Behavioral Assessment of the Dysexecutive Syndrome for Children</i>
BASC-2	<i>The Behavior Assessment System for Children, Second Edition</i>
BDEFS-CA	<i>Barkley Deficits in Executive Functioning Scale- Children and Adolescents</i>
BRIEF	<i>Behavior Rating Inventory of Executive Function</i>
BRIEF-2	<i>Behavior Rating Inventory of Executive Function, Second Edition</i>
BRIEF-P	<i>Behavior Rating Inventory of Executive Function, Preschool</i>
BRIEF-SR	<i>Behavior Rating Inventory of Executive Function, Self-Report</i>
CCI	Coeficiente de Correlação Interclasse
CEFI	<i>Comprehensive Executive Function Inventory</i>
CEFS	<i>Children's Executive Functions Scale</i>
CBQ	<i>Children's Behavior Questionnaire</i>
CBQ-SE	<i>Children's Behavior Questionnaire – Short Form</i>
CBQ-VSE	<i>Children's Behavior Questionnaire – Very Short Form</i>
COSMIN	<i>COnsensus-based Standards for the selection of health Measurement INstruments</i>
CHECK	<i>Child Evaluation Checklist</i>
CHEXI	<i>The Childhood Executive Functioning Inventory</i>
DEX-C	<i>Dysexecutive Questionnaire for Children</i>
D-REF	<i>Delis Rating of Executive Function</i>
EFS	<i>Executive Functioning Scale for Children</i>
EFORTS	<i>Executive Function and Occupational Routines Scale</i>
EUA	Estados Unidos da América
FE	Funções Executivas
IVC	Índice de Validade de Conteúdo
LEAF	<i>Learning, Executive, and Attention Functioning</i>
PEA	Perturbação do Espectro do Autismo
PHDA	Perturbação de Hiperatividade e Défice de Atenção
POC	Perturbação Obsessivo-Compulsiva
PRISMA	<i>Preferred Reporting Items for Systematic Review and Meta-Analyses</i>

QEF	<i>Questionnaire Executive Function</i>
RAYYAN	<i>Intelligent Systematic Review</i>
REEF	<i>Ratings os Everyday Executive Functioning</i>
TSI	<i>Thinking Skills Inventory</i>

1. Introdução Teórica

A descrição do caso de Phineas Gage por Harlow, ainda no século XIX, foi a primeira tentativa de descrição do comprometimento das Funções Executivas (FE) decorrente de uma lesão dos lobos frontais (Gazzaniga et al., 1998). Phineas Gage, sobreviveu a um acidente de trabalho no qual sofreu uma lesão grave no seu cérebro: uma barra de ferro atravessou o seu crânio, destruindo a maior parte do seu lobo frontal esquerdo. Phineas Gage originou assim um dos maiores estudos de caso associado às FE, uma vez que após o acidente várias mudanças de comportamento e personalidade começaram a ser perceptíveis. Devido a este e a outros casos, diversos investigadores começaram a ter curiosidade em investigar o papel dos lobos frontais e o conceito de FE (Goldstein & Naglieri, 2014).

Até aos dias de hoje não existe consenso no que diz respeito à definição de FE, devido à sua complexidade e pela sua conceptualização multidimensional (Anderson et al., 2001). As FE podem ser definidas como um conjunto abrangente de processos cognitivos que estão inter-relacionados, sendo que estas competências, para além de multifacetadas, dizem respeito a habilidades necessárias para comportamentos direcionados a um objetivo (Anderson et al., 2001). É, portanto, um amplo processo cognitivo usado para direcionar o comportamento, especificamente em situações onde algumas respostas devem ser inibidas e outras precisam ser iniciadas (Goldstein & Naglieri, 2014). As FE requerem ainda a capacidade de responder de forma flexível e adaptativa às mudanças ambientais em busca de objetivos de longo prazo. São também consideradas funções mentais responsáveis pela capacidade de autorregulação e autogerenciamento de uma diversidade de processos cognitivos (Thompson & Steinbeis, 2020).

Embora não exista um acordo quanto às suas componentes, os subdomínios executivos, mais frequentemente referidos pelos autores, incluem a inibição, o raciocínio abstrato, a regulação emocional, a automonitorização, a iniciativa, a memória de trabalho, a flexibilidade cognitiva, o planeamento e organização, a resolução de problemas e a atenção (Zelazo & Muller, 2002; Goldstein & Naglieri, 2014). Logo, as FE incorporam uma série de habilidades e modalidades cognitivas que afetam os aspetos do comportamento, incluindo a habilidade social, fundamentais para um funcionamento eficiente na vida quotidiana (Anderson et al., 2001).

Quando falamos em FE, devemos distinguir as FE frias das FE quentes, uma vez que a resolução de problemas de vida diária, envolve a combinação destas (Zelazo & Muller, 2002). As FE frias estão relacionadas com processos cognitivos ativados perante problemas abstratos e descontextualizados e habilidades orientadas para o futuro. Estas FE são dirigidas a objetivos, que não se manifestam sob contextos emocionais e analíticos (Miyake et al., 2000).

Por sua vez, as FE quentes relacionam-se com aspetos afetivos requeridos em situações significativas que envolvem a regulação do afeto e da motivação, direcionados a objetivos e são processos cognitivos orientados para o futuro, eliciados em contextos que geram emoção e motivação (Goldstein & Naglieri, 2014).

1.1. Componentes Executivas

Embora também não exista um consenso quanto às dimensões que compõem as FE, na literatura são referidos com frequência três domínios centrais, nomeadamente o controlo inibitório, a memória de trabalho e a flexibilidade cognitiva que assumem uma importância fulcral em diversos modelos teóricos (Diamond, 2012). Habilidades como a memória, inibição e a flexibilidade são universalmente aceites como habilidades essenciais das FE porque são necessárias para atividades focadas e direcionadas a objetivos (Miyake et al., 2000).

O controlo inibitório consiste na capacidade de gerir a atenção, o comportamento, as emoções e os pensamentos para superar distrações internas e externas e executar o que é importante num determinado momento (Diamond, 2012). É a FE essencial para controlar o comportamento de uma pessoa, substituindo respostas habituais ou exercendo autocontrolo, como por exemplo, resistir à tentação de não completar uma tarefa. Logo, é a capacidade de manter e/ou parar uma ação (Diamond, 2012), inibindo comportamentos, filtrando pensamentos, controlando impulsos e resistindo a tentações, estando também responsável por levar o sujeito a pensar antes de agir e inibir deliberadamente estímulos que não são relevantes para a tarefa que está a ser executada (Miyake et al., 2000; Goldstein & Naglieri, 2014).

Segundo Goldstein e Naglieri (2014), a memória de trabalho é o sistema cerebral que fornece, temporariamente, o armazenamento e manipulação de informações e está envolvida

em tarefas cognitivas como a compreensão da linguagem, aprendizagem e raciocínio e é o processo que manipula ativamente informações fonológicas ou visuoespaciais mentais (Baddeley & Hitch, 1974). Para além de manter e manipular informações, esta FE possibilita relacionar e integrar informações, lembrar sequências ou ordens e projetar sequências de ações no futuro (Diamond, 2013). Resumidamente, é a FE que permite manter a informação disponível, manipulá-la, chegando assim à solução do problema (Miyake et al., 2000).

Por sua vez, a flexibilidade cognitiva, está relacionada com a capacidade de resolver problemas, de fazer transições e até de mudar perspectivas (Spitzer et al., 2019). Segundo Diamond (2013), é também a capacidade de mudar o foco atencional, alternando entre tarefas e focos, bem como permite que o sujeito considere diferentes perspectivas. Possibilita que este se adapte às demandas do meio ambiente, que possua habilidades intimamente ligadas à criatividade, mas também viabiliza a resolução de problemas (Diamond, 2013). Permite então que o ser humano altere cognições e comportamentos de acordo com as mudanças ambientais (Lezak et al., 2004).

Segundo alguns autores, a partir destes três domínios principais, desenvolvem-se outros mais complexos, como o planeamento, a resolução de problemas, a organização e a fluência verbal (Korzeniowski & Ison, 2019). Outros domínios principais das FE incluem a atenção controlada, o processamento sequencial sustentado, o planeamento e execução de comportamentos direcionados a metas, a resolução de problemas e a organização, necessários para iniciar e concluir propósitos e foram validados como comportamentos de FE (Castellanos et al., 2016).

1.2. Modelos Teóricos das FE

Um dos modelos teóricos mais frequentemente citado na literatura é o Modelo dos Três Fatores de Miyake et al. (2000) destacando três componentes das FE, o *shifting*, o *updating* e a inibição. Este modelo também suporta a ideia de que a inibição, a memória de trabalho e a flexibilidade cognitiva, apesar de distintos, não são completamente independentes (Miyake et al., 2000).

Outro modelo destacado com frequência é o Modelo de Desenvolvimento das FE de Diamond (2013), que interpreta as FE como sendo um conjunto de processos mentais *top-down* e defende que a memória de trabalho, o controlo inibitório e a flexibilidade cognitiva

são componentes dissociados, e com diferentes trajetórias de desenvolvimento. Porém defende igualmente que estas estão também relacionadas, uma vez que a partir destas componentes principais, desenvolvem-se FE de ordem superior, como o raciocínio, resolução de problemas e o planeamento (Diamond, 2013).

Todavia, outros autores têm também destacado perspectivas diferentes. Por exemplo, em 2002, Anderson descreve numerosas componentes das FE, considerando que os principais elementos dos processos associados às FE incluíam a antecipação, a seleção de metas, o planeamento, o início da atividade, a autorregulação, a flexibilidade mental, a atenção e o uso de *feedback* (Anderson, 2002).

Mais recentemente, Levine e Craik (2012), salientam as FE de ordem superior necessárias para executar tarefas complexas no ambiente, incluindo a seleção e perceção, manutenção, a memória de trabalho, o comportamento autodirigido, o planeamento e organização, a regulação e controlo comportamental e a tomada de decisão (Levine & Craik, 2012; Goldstein & Naglieri, 2014).

Por fim, para Goldstein, Naglieri, Princiotta e Otero (2014), as FE são um termo usado para uma diversidade de hipóteses cognitivas, processos esses realizados por áreas pré-frontais dos lobos frontais, e incluem o planeamento, a memória de trabalho, a atenção, a inibição, o automonitoramento, a autorregulação e a iniciação (Goldstein et al., 2014). Importa referir que estes modelos teóricos das FE não são exaustivos, pelo que apenas são alguns exemplos de modelos que constatarem os diversos processos cognitivos das FE.

1.3. Desenvolvimento das FE

Posto isto, é essencial compreender o desenvolvimento das FE, e que um desenvolvimento saudável destes processos cognitivos, é fundamental para resultados positivos ao nível dos vários contextos de vida diária (Halse et al., 2019).

O desenvolvimento inicial das FE está associado a vários fatores de interação, alguns que são internos ao próprio indivíduo em desenvolvimento (fatores individuais) e outros que existem no meio ambiente (fatores ambientais) (Thompson & Steinbeis, 2020). Além disso, as FE são um aspeto central da cognição que passa por mudanças importantes e que são fundamentais para a infância. Estas mudanças estão relacionadas com o desempenho académico, satisfação geral com a vida, e até mesmo em termos de saúde física (Nesbitt &

Farran, 2021). Assim sendo, as FE são um importante tópico de estudo, principalmente neste estágio de desenvolvimento, uma vez que é durante este período que as FE começam a surgir e se vão desenvolvendo (Nesbitt & Farran, 2021).

As FE surgem então já no início do primeiro ano de vida, desenvolvendo-se posteriormente, mais precisamente, entre os 2 e os 5 anos de idade, constituindo-se a primeira infância como um período inicial de rápida mudança no desenvolvimento (Camerota et al., 2016). Todavia, um desempenho ao nível adulto em muitos testes de avaliação das FE são alcançados por volta dos 12 anos de idade (Zelazo & Muller, 2002). Contudo, diversos autores referem que o desempenho executivo continua a maturar, posteriormente, até à vida adulta (Zelazo & Muller, 2002).

Sendo assim, a idade escolar constitui-se um período crítico para o desenvolvimento das FE, uma vez que as crianças estão sujeitas a participar em atividades mais estruturadas, exigindo, dessa forma, mais das suas FE. Isto porque as FE são utilizadas numa variedade de atividades, como por exemplo, para iniciar ações, mudar de tarefas, inibir impulsos, focar numa tarefa específica ou até mesmo para a interação social (Halse et. al., 2019).

Logo, as FE desempenham um papel importante nas crianças, ao nível das habilidades académicas, sociais, emocionais e comportamentais (Howard et al., 2015). E uma trajetória evolutiva adequada das mesmas durante a infância prediz uma maior qualidade de vida e menor incidência de perturbações de conduta na adolescência e na idade adulta (Korzeniowski & Ison, 2019).

As FE quando bem desenvolvidas permitem que as crianças gerenciem tarefas adequadas à idade de forma mais independente. No entanto, um atraso ou limitações no seu desenvolvimento podem causar graves problemas funcionais (Frisch & Roseblum, 2014). Dessa forma, alterações nas FE são um problema de saúde clínico significativo não apenas devido à sua prevalência em lesões e distúrbios, mas também devido à influência de alterações nas FE prejudiciais ao desempenho académico, podendo resultar em problemas de memória, fracasso educacional e dificuldades de aprendizagem (Castellanos et al., 2016). Segundo Karr e Garcia-Barrera (2016), um desenvolvimento inadequado das FE está também relacionado a dificuldades na adaptação do indivíduo ao ambiente (Karr & Garcia-Barrera, 2016). Além do mais, crianças com défices nas FE podem vivenciar desafios diários consideráveis na gestão das suas rotinas, como agir com lentidão e esquecimentos (Frisch & Roseblum, 2014). Segundo Anderson et al. (2001), alterações no desenvolvimento das FE,

podem conduzir à falta de atenção e planeamento, dificuldades na geração e implementação de estratégias, incapacidade de utilizar *feedback* e inflexibilidade de pensamento (Anderson et al., 2001).

Assim, pode-se concluir que, potenciar o desenvolvimento das FE nos primeiros anos de vida é fundamental, pois problemas executivos na primeira infância predizem posteriores dificuldades, e défices nas FE geralmente não desaparecem, podendo mesmo aumentar ao longo do tempo (Diamond, 2012). Mais ainda, o papel crítico das FE no desenvolvimento, é frequentemente observável numa variedade de perturbações de desenvolvimento (Howard et al., 2015).

Por causa do papel significativo das FE para a aprendizagem, memória e resultados académicos, a avaliação das alterações executivas que podem afetar significativamente o sucesso escolar é uma componente comum de avaliação neuropsicológica em crianças (Castellanos et al., 2016). Para além do mais, a identificação de alterações executivas, é importante para o planeamento de uma intervenção eficaz (Wang et al., 2018). Isto sugere a importância de avaliar tanto a gestão das rotinas diárias das crianças, assim como as suas FE, utilizando instrumentos adequados para esse fim (Frisch & Roseblum, 2014). No entanto, identificar alterações nas trajetórias de desenvolvimento das FE, com vista à implementação de intervenções precoces e efetivas, é um dos desafios atuais da neuropsicologia infantil (Korzeniowski & Ison, 2019).

1.4. Alterações nas FE

Quando ocorrem alterações ao nível das FE, a chamada síndrome disexecutiva, sucedem-se alterações cognitivo-comportamentais, que podem manifestar-se de diversas formas, incluindo dificuldades na seleção da informação, dificuldades na tomada de decisão, problemas na organização, comportamento estereotipado, comportamento impulsivo, dificuldades em planear com antecedência e dificuldades em adaptar o comportamento a mudanças (Huizinga & Smidts, 2011). São alguns exemplos que ditam uma série de problemas na vida diária (Dias et al., 2010). Segundo Halvorsen et. al (2018), este conceito é a característica chave de várias perturbações do neurodesenvolvimento, mais notavelmente na Perturbação de Hiperatividade e Défice de Atenção (PHDA), na Perturbação do Espectro do Autismo (PEA), nas Perturbações de Conduta, no Comportamento Antissocial, na Perturbação Obsessivo-Compulsiva (POC), na Perturbação de Aprendizagem (e.g. dislexia

e discalculia), mas também em quadros neurológicos, como a Epilepsia (Peters et al., 2012; Bonillo et al., 2012).

1.5. Avaliação das FE

Por tudo isto, é fundamental a avaliação das FE em crianças, sendo que atualmente existem distintas formas de avaliação (Santana et al., 2019). No entanto, apesar do crescente interesse pelas FE na pré-escola e em crianças em idade escolar, ainda é incomum avaliar as FE em crianças menores de seis anos de idade, mesmo em grupos de alto risco, como por exemplo, os nascidos prematuros. Isto deve-se, em parte, à escassez de ferramentas adequadas para avaliar as FE em crianças pequenas (O'Mehager et al., 2019).

Não obstante, alterações executivas são muito difíceis de detetar no ambiente clínico, utilizando habituais ferramentas de avaliação padronizadas, como por exemplo testes de desempenho (Anderson et al., 2001). Assim, apesar da avaliação das FE ser essencial devido à sua importância para um desenvolvimento típico, para adquirir habilidades de vida favoráveis e para uma vida bem-sucedida e adaptada, não existe consenso relativamente a um protocolo de avaliação, pelo que existem vários instrumentos de avaliação das FE, como testes, escalas, baterias e questionários/inventários (Anderson et al., 2001; Santana et al., 2019).

As formas mais comuns de avaliação das FE nas crianças são os testes baseados em desempenho, entrevistas com pais e professores e escalas de avaliação padronizadas preenchidas pelos mesmos (O'Mehager et al., 2019). Dado que é demorado e dispendioso avaliar crianças usando medidas baseadas em desempenho (e.g. teste de trihas, teste de stroop), os questionários são um método de triagem comum usado para avaliar comportamentos problemáticos e o funcionamento infantil (Gutierrez et al., 2021). Os questionários têm sido bastante úteis na identificação da disfunção executiva e são, frequentemente, empregados para obter impressões dos pais, professores ou outros cuidadores do indivíduo em várias situações do dia-a-dia (Schneider et al., 2019). Segundo Tamm e Peugh (2018), estes instrumentos foram desenvolvidos de forma a reunir as observações do avaliador, dos comportamentos diários associados às FE das crianças, refletindo a aplicação real das habilidades das FE pelas crianças nas tarefas diárias (Shum et al., 2020).

Segundo Kihlstrom (2021), os questionários de avaliação para além de serem considerados o método com maior validade ecológica, ou seja, informação válida que possa ser generalizada para as diferentes populações, situações e variáveis, são instrumentos de relato do funcionamento diário num ambiente natural descrito pelas pessoas que interagem com o indivíduo diariamente. Além de serem mais reflexivos no que toca ao quotidiano e ao funcionamento executivo, os questionários de avaliação das FE mostram associações mais fortes com o lobo frontal, centro das FE e o maior responsável pelos nossos atos cognitivos (Schneider et al., 2019).

Para além disto, os questionários avaliam diferentes domínios executivos, e as pontuações geralmente apresentam maior confiabilidade relativamente a outras escalas (Karr & Garcia-Barrera, 2016). Também permitem recolher informação dos vários contextos ao redor da criança, aspeto fundamental para determinar o nível de comprometimento das FE (Schneider et al., 2019). Por exemplo, a recolha de dados dos educadores é fundamental, principalmente para crianças em idade pré-escolar, uma vez que os comportamentos pré-escolares podem ser mais fáceis de julgar, especialmente no contexto de sala de aula mais estruturado. Os professores têm um grupo de comparação composto por outras crianças e, portanto, podem descrever com mais facilidade os comportamentos que se afastam da norma (Tamm & Peugh, 2018). Sendo assim, os questionários são considerados a melhor abordagem para a avaliação das FE e a mais recomendada pelos pesquisadores (Karr & Garcia-Barrera, 2016).

A título de exemplo, e segundo Castellanos et al. (2016), domínios tais como a organização, planeamento, memória de trabalho, flexibilidade cognitiva, controlo da atenção e habilidades de inibição, por exemplo, são rotineiramente avaliados usando medidas de FE baseadas em questionários, como o *Behavior Rating Inventory of Executive Function* (BRIEF) (Esbensen et al., 2019), *Barkley Deficits in Executive Functioning Scale—Children and Adolescents* (BDEFS-CA) (Mashhadi et al., 2020) e *Comprehensive Executive Function Inventory* (CEFI) (Krieger et al., 2019).

O *Behavior Rating Inventory of Executive Function* (BRIEF) é o questionário de FE mais comumente usado (Schneider et al., 2019). É um instrumento de classificação que pode ser preenchido pelos pais e professores de forma a auxiliar na identificação de dificuldades de FE em crianças dos 3 aos 5 anos de idade. É um instrumento para ser utilizado por vários informantes e permite que estes respondam exatamente aos mesmos itens, sendo que essa

consistência permite uma investigação de alta confiabilidade, dentro de uma faixa etária na qual o auto-relato é normalmente inatingível (Schneider et al., 2019). Este formulário é então direcionado a pais e professores e ainda contém uma lista de verificação de comportamentos de auto-relato para crianças e adolescentes e avalia áreas, tais como, a inibição, automonitoramento, mudança, controlo emocional, iniciação, conclusão da tarefa, memória de trabalho, planeamento/organização, monitoramento de tarefas e organização de materiais (Castellanos et al., 2016).

Por exemplo, os professores ao preencherem o *Behaviour Rating Inventory of Executive Function, Preschool* (BRIEF-P), estão a avaliar os comportamentos de FE em ambiente escolar produzindo, dessa forma, um *score* total de acordo com a idade dos vários índices de FE. No total, os valores mais altos vão indicar um desempenho inferior. Este questionário serve para avaliar crianças dos 2 aos 5 anos de idade e é uma escala com confiabilidade e validade adequadas (α 0,956), daí ser um instrumento de avaliação frequentemente usado para avaliar as FE nas crianças (Tamm & Peugh, 2019).

Para além deste, existem outros questionários, como por exemplo, o *Comprehensive Executive Function Inventory* (CEFI) que inclui 100 itens e mede domínios das FE como a atenção, regulação emocional, flexibilidade cognitiva, controlo inibitório, iniciação, organização, planeamento, automonitoramento e memória de trabalho (Castellanos et al., 2016). Segundo Krieger et al. (2019), este instrumento apresenta confiabilidade muito boa (0,99). Já o *Barkley Deficits in Executive Functioning Scale- Children and Adolescents* (BDEFS-CA) inclui 70 itens e mede as FE em atividades de vida diária, como gerir o tempo, organização e resolução de problemas, autocontrolo, automotivação e autoregulação de emoções (Castellanos et al., 2016). A consistência interna e a validade são boas (.91) (Kamradt et al., 2019). Estes são apenas alguns exemplos de questionários que avaliam as FE em crianças.

1.6. Objetivos de Investigação

Apesar da relevância das FE no desenvolvimento típico e atípico e da necessidade da sua avaliação, esta informação não se encontra sistematizada. Assim, o presente estudo tem como objetivo realizar uma revisão sistemática da literatura existente sobre a matéria, versando em particular em questionários/inventários desenvolvidos para este fim. Este

estudo contribuía assim para sistematizar os questionários/inventários mais comumente usados para a avaliação das FE em crianças e analisar as suas características psicométricas.

2. Método

2.1 Pesquisa

2.1.1 Questão de Investigação

Quais são os questionários/inventários mais utilizados para a avaliação das FE em crianças até aos 12 anos e respetiva validade psicométrica?

2.1.2 Base de Dados

A pesquisa de artigos foi realizada a partir de Outubro de 2021, através da utilização do motor de busca *EBSCO*, nas bases de dados de Psicologia, nomeadamente, *Psychology and Behavioral Sciences*, *Academic Search Complete*, *APA PsycInfo* e *Eric*.

2.1.3 Termos de Pesquisa

A pesquisa foi realizada recorrendo às seguintes palavras-chave: *Executive Functioning OR Executive Function; AND Questionnaire OR Survey OR Scale OR Instrument OR Inventory AND Psychometric Properties OR Validity OR Reliability*. Foi adicionado também o filtro *Childhood (birth – 12 years)*.

2.1.4 Critérios de Elegibilidade

Foram estabelecidos alguns critérios de inclusão, de forma a que os resultados da pesquisa fossem de encontro à temática que se pretende estudar, sendo estes: 1) questionários/inventários de funções executivas; 2) questionários/inventários dirigidos a crianças, pais e/ou professores; 3) questionários/inventários aplicados a crianças até aos 12 anos de idade.

Relativamente aos critérios de exclusão, foram os seguintes: 1) questionários/inventários não direcionados para as funções executivas; 2) estudos não publicados em inglês, português ou espanhol; 3) dissertações, revisões da literatura e livros.

A pesquisa dos artigos científicos foi compreendida entre os anos de 2011 e 2021, com o auxílio do programa *Rayyan (Intelligent Systematic Review)*; Ouzzani et al., 2016).

2.2 Procedimentos

2.2.1 Recolha de Dados

A seleção dos artigos foi realizada respeitando os critérios de inclusão/exclusão estabelecidos, sendo que a primeira fase da seleção dos artigos teve por base a análise do título e o resumo dos mesmos. Relativamente à segunda fase, esta exigiu uma leitura atenta e integral de cada artigo, já previamente selecionados. Ambas as fases de análise foram realizadas por duas pessoas, de forma individual e independente, em que as discórdias foram resolvidas através de uma discussão até haver consenso mútuo nas decisões. Esta revisão sistemática segue também as diretrizes *PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)*, de forma a melhorar a qualidade do relatório (Page et al., 2021).

2.3 Estratégia de análise de dados

Após a seleção dos estudos, foi realizada a análise e o resumo dos mesmos, dando ênfase aos resultados relacionados com os questionários/inventários de funções executivas em crianças.

Foi igualmente realizada uma avaliação do risco de viés, que se refere à avaliação da validade interna e externa dos estudos incluídos numa revisão sistemática, que constitui uma parte fundamental do processo (McGuinness & Higgins, 2020). Esta avaliação foi realizada tendo como referência a ferramenta *COSMIN checklist (CONsensus-based Standards for the selection of health Measurement Instruments)*, que tem por objetivo desenvolver uma lista de verificação de padrões que avaliam a qualidade metodológica dos estudos sobre propriedades de medida, melhorando a seleção dos instrumentos de medição (Mokking et al., 2018). Esta ferramenta propõe a aplicação de nove propriedades de medida a cada estudo,

sendo eles Consistência Interna, Confiabilidade, Erro de Medição, Validade do Conteúdo, Validade Estrutural, Teste de Hipóteses, Validade Transcultural, Validade do Critério e Capacidade de Resposta, em que cada uma é classificada de “muito bom” a “inadequado” resultando numa pontuação para cada propriedade de medida (Mokking et al., 2018).

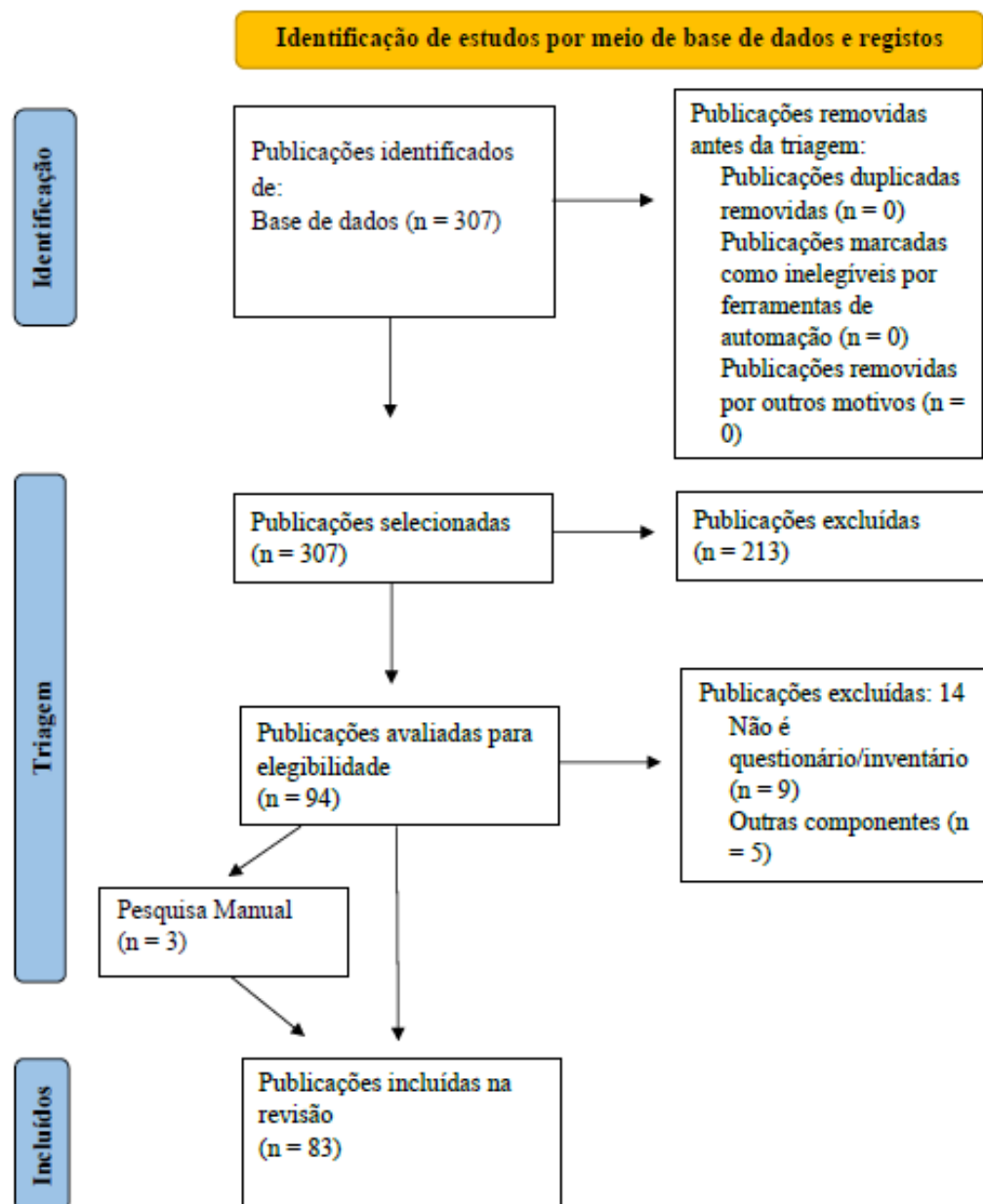
Todavia, para este estudo, foi determinado avaliar apenas quatro domínios, ou seja, a Validade, a Confiabilidade, a Consistência Interna e a Capacidade de Resposta. Isto porque a taxonomia do *COSMIN* compreende três domínios principais, nomeadamente, a Confiabilidade, a Validade e a Capacidade de Resposta (Cf. Anexo 1) (Mokking et al., 2010). Relativamente à Consistência Interna, foi igualmente avaliada, uma vez ser uma propriedade de medida frequentemente avaliada pelos pesquisadores nos seus estudos.

3. Resultados

Através de buscas nas bases de dados eletrônicas, um total de 307 artigos foram inicialmente identificados, não havendo duplicados. Posteriormente, através de uma pesquisa manual, foram ainda incluídos mais 3 estudos por respeitarem todos os critérios de inclusão previamente estabelecidos. Procedeu-se a uma primeira leitura, analisando apenas os títulos e resumos dos artigos, sendo que 213 foram excluídos. Após a primeira fase de pesquisa, procedeu-se à leitura integral de um total de 94 artigos para verificação dos critérios de inclusão/exclusão. Após esta leitura, 14 estudos foram também excluídos. Dessa forma, seguiram para inclusão no estudo 83 artigos, por respeitarem todos os critérios definidos. Esta seleção encontra-se representada no Fluxograma PRISMA na figura 1.

Figura 1

Fluxograma PRISMA 2020



Uma vez que um número elevado de estudos utilizaram o mesmo questionário, nomeadamente o *Behavior Rating Inventory of Executive Function (BRIEF)*, foram analisados apenas os artigos que validaram ou traduziram este instrumento, sendo por analisados um total de 41 estudos.

3.1. Características Gerais dos Estudos

Os estudos foram realizados em diversos países, tais como: Índia, EUA, Espanha, Israel, Brasil, Irão e Canadá. Todavia, a maioria dos estudos incluídos foram realizados nos EUA e Espanha, sendo que nenhum foi realizado em Portugal. Apenas foi possível identificar um único artigo de validação do questionário *Behavior Rating Inventory of Executive Function (BRIEF)* para português do Brasil.

Os estudos foram publicados entre os anos de 2011 a 2021, no entanto, os anos em que houve mais estudos foram 2014, 2015, 2016 e 2019.

3.2. Questionários/Inventários de FE

Os estudos incluídos nesta revisão sistemática permitiram identificar um total de 18 questionários/inventários para avaliação das FE em crianças. Na Tabela 1, estão apresentados os diferentes questionários/inventários utilizados. Em particular, o *Behavior Rating Inventory of Executive Function (BRIEF)* foi aplicado em 56 artigos, o *Behavior Rating Inventory of Executive Function, Preschool (BRIEF-P)* em 11 artigos, o *Behavior Assessment System for Children, Second Edition (BASC-2)* em 8 artigos e o *Childhood Executive Functioning Inventory (CHEXI)* em 6 dos artigos incluídos. Os restantes instrumentos apresentam-se num número reduzido de estudos, tal como podemos observar na Tabela 1.

Tabela 1

Visão Geral dos Instrumentos

Autor	Local	Ano	Instrumento	Número de Artigos^a
Byerley & Donders	EUA	2013	<i>BRIEF-SR (Behavior Rating Inventory of Executive Function-Self-Report)</i>	1
Castellanos et. al.	Índia	2016	<i>LEAF (Learning, Executive, and Attention Functioning)</i>	1
Catale et. al.	Bélgica/Suécia	2015	<i>CHEXI (Childhood Executive Functioning Inventory)</i>	6
Dowdy et. al.	EUA	2011	<i>BASC-2 (The Behavior Assessment System for Children, Second Edition)</i>	8
Esbensen et. al.	EUA	2019	<i>BRIEF (Behaviour Rating Inventory of Executive Function)</i>	56
Ezpeleta et. al.	Espanha	2015	<i>BRIEF-P (Behaviour Rating Inventory of Executive Function, Preschool Version)</i>	11
Frisch & Rosenblum	Israel	2014	<i>EFORTS (Executive Function and Occupational Routines Scale)</i>	1
FY Siu & Zhou	Brasil	2014	<i>BADS-C (Behavioral Assessment of the Dysexecutive Syndrome for Children)</i>	1
Guerten et. al.	Índia	2016	<i>QEF (Questionnaire of Executive Functioning)</i>	1
Jiménez & Lucas-Molina	Espanha	2018	<i>BRIEF-2 (Behavior Rating Inventory of Executive Function, Second Edition)</i>	4
Krieger et. al.	Espanha	2019	<i>CEFI (Comprehensive Executive Function Inventory)</i>	1
Korzeniowski et. al.	Espanha	2019	<i>EFS (Executive Functioning Scale for Children)</i>	1
Mashhadi et. al.	Irão	2020	<i>BDEFS-CA (Barkley Deficits in Executive Functioning Scale- Children and Adolescents)</i>	2
Nilsen et. al.	Canadá	2017	<i>REEF (Ratings of Everyday Executive Functioning)</i>	1
Osa et. al.	Espanha	2014	<i>CBQ (Children's Behavior Questionnaire)</i>	1
Rosenblum et. al.	Israel	2019	<i>CHECK (Child Evaluation Checklist)</i>	1
Wang et. al.	EUA	2018	<i>TSI (Thinking Skills Inventory)</i>	1
Warnick et. al.	EUA	2015	<i>D-REF (Delis Rating of Executive Function)</i>	1

Nota. ^aNúmero de artigos encontrados que utilizam os respectivos instrumentos

Através da leitura integral dos artigos incluídos no estudo, foi possível identificar mais instrumentos. Um deles é o *Children's Executive Functions Scale (CEFS)*, com 99 itens, projetado para medir uma variedade de comportamentos relacionados ao funcionamento executivo. Outro questionário identificado foi o *Dysexecutive Questionnaire for Children (DEX-C)*, instrumento de 20 itens elaborado para avaliar aspectos cognitivos, comportamentais, emocionais e problemas motivacionais associados à disfunção executiva. Por fim, foi perceptível a existência de outras versões do *Child Behavior Questionnaire (CBQ)*, o *Child Behavior Questionnaire, Short Form (CBQ-SE)*, com 94 itens e o *Child Behavior Questionnaire Very Short Form (CBQ-VSE)*, com 36 itens.

3.2.1. Características dos instrumentos

Estes instrumentos apresentam características díspares, avaliam diferentes aspectos, mas também possuem objetivos distintos, embora o principal objetivo em comum seja a avaliação das FE em crianças. Alguns dos questionários, como é o caso do CBQ e do TSI, embora não sejam elaborados com o intuito de avaliar as FE, avaliam algumas componentes executivas. Dessa forma, os instrumentos avaliam diferentes processos cognitivos e os comportamentos relacionados ao funcionamento executivo, analisando défices de habilidades, por meio de rotinas diárias.

Estes instrumentos avaliam sobretudo, crianças e adolescentes, dos 2 aos 21 anos. E, apesar de a população-alvo deste estudo ser crianças até aos 12 anos, pode-se concluir que os questionários apresentam uma grande diversidade de faixas etárias, pelo que é um ponto positivo. Contudo, foi perceptível que há ainda poucos questionários dirigidos a crianças em idade pré-escolar.

Outra curiosidade, é que os questionários apresentam um número díspar de itens. O CHEXI é o que contém menos, cerca de 26, e o CBQ o que contém mais (195), seguido do CEFI, com 100 itens. Por fim, é notório que a grande maioria dos questionários é dirigido a pais e professores (Cf. Tabela 2).

Tabela 2

Características dos Instrumentos

Instrumento	Objetivo	O que avalia	Número de Itens	Faixa Etária	Preenchimento
<i>BADS-C (Behavioral Assessment Dysexecutive Syndrome for Children)</i>	Diagnostica défices no funcionamento executivo das crianças	Flexibilidade, Resolução de Problemas, Impulsividade, Planeamento e Incapacidade de usar o <i>Feedback</i>	86 Itens	8 aos 15 anos	Pais
<i>BASC-2 (Behavior Assessment System for Children, Second Edition)</i>	Ajuda no diagnóstico e tratamento de défices de FE de crianças, adolescentes e jovens adultos	Resolução de Problemas, Controlo da Atenção, Controlo do Comportamento e Controlo Emocional	30 Itens	6 aos 21 anos	Pais, professores e auto-relato
<i>BDEFS-CA (Barkley Deficits in Executive Functioning Scale – Children and Adolescents)</i>	Fornece avaliação de défices de funcionamento executivo nas atividades diárias	Atenção, Inibição, Memória de Trabalho, Planeamento e Resolução de Problemas	70 Itens	6 aos 17 anos	Pais
<i>BRIEF (Behaviour Rating Inventory of Executive Function)</i>	Avalia o comportamento das FE em casa e no ambiente escolar	Inibição, Memória de Trabalho, Planeamento e Organização	86 Itens	5 aos 18 anos	Pais e Professores
<i>BRIEF-2 (Behaviour Rating Inventory of Executive Function, Second Edition)</i>	Fornece as informações necessárias para ajudar crianças e adolescentes com disfunção executiva	Inibição, Mudança, Automonitoramento, Controlo Emocional, Iniciação, Memória de Trabalho, Planeamento e Organização	63 Itens	5 aos 18 anos	Pais e professores
<i>BRIEF-P (Behaviour Rating Inventory of Executive Function, Preschool Version)</i>	Projetado para medir possíveis dificuldades de FE em crianças em idade pré-escolar	Memória de trabalho, Planeamento, Organização e Inibição	63 Itens	2 aos 5 anos	Pais e professores

<i>BRIEF-SR (Behaviour Rating Inventory of Executive Function, Self-Report)</i>	Capturar a visão da criança/adolescente do seu próprio comportamento	Inibição, Mudança, Controlo Emocional, Memória de Trabalho, Planeamento/Organização, Conclusão de Tarefa	80 Itens	11 aos 18 anos	Auto-relato
<i>CBQ (Children's Behavior Questionnaire)</i>	Serve para medir o temperamento das crianças	Antecipação Positiva, Sorriso/Risadas, Prazer de alta e baixa intensidade, Nível de Atividade, Impulsividade, Timidez, Desconforto, Medo, Raiva/Frustração, Tristeza, Tranquilidade, Controlo Inibitório, Foco Atencional e Sensibilidade Perceptiva	195 Itens	3 aos 7 anos	Cuidadores
<i>CEFI (Comprehensive Executive Function Inventory)</i>	Serve para classificar comportamentos dos pontos fortes e fracos das FE	Atenção, Regulação Emocional, Flexibilidade, Controlo Inibitório, Planeamento, Automonitoramento e Memória de Trabalho	100 Itens	5 aos 18 anos	Pais, professores e auto-relato
<i>CHECK (Child Evaluation Checklist)</i>	Destina-se a identificar risco para condições invisíveis do neurodesenvolvimento, concentrando-se na atividade diária das crianças	Controlo Motor, Habilidades Verbais, Autoregulação e Habilidades Sociais, nomeadamente Memória de Trabalho, Autoregulação do Afeto, Motivação e Excitação	40 Itens	3 aos 6 anos	Pais
<i>CHEXI (Childhood Executive Functioning Inventory)</i>	Serve para medir o funcionamento executivo em crianças.	Memória de trabalho, Planeamento, Inibição e Regulação	26 Itens	5 aos 11 anos	Pais, professores ou pessoa próxima
<i>D-REF (Delis Rating of Executive Function)</i>	Permite relatar os comportamentos observados que identificam problemas de FE em crianças e adolescentes	Atenção, Memória de Trabalho, Controlo de Impulso, Conformidade, Controlo da Raiva e Raciocínio Conceptual	36 Itens	5 aos 18 anos	Pais, professores e auto-relato

<i>EFORTS (Executive Functioning and Occupational Routines Scale)</i>	Examina o controlo executivo das crianças nas rotinas diárias ocupacionais	Inibição, Memória de Trabalho, Autoregulação, Resolução de Problemas e Planeamento	32 Itens	3 aos 10 anos	Pais
<i>EFS (Executive Functioning Scale for Children)</i>	Permite identificar a frequência com que a criança apresenta uma série de comportamentos que são indicativos de possíveis défices de FE	Impulsividade, Hiperatividade, Controlo Emocional, Memória Funcional, Capacidade de Planificação, Capacidade de Organização, Flexibilidade, Capacidade de Concentração e Foco de Atenção	30 Itens	6 aos 10 anos	Pais/cuidadores
<i>LEAF (Learning Executive, and Attention Functioning)</i>	Mede amplamente as habilidades neuropsicológicas do funcionamento executivo e aprendizagem	Atenção, Velocidade de Processamento, Concentração, Organização Visuo-Espacial, Planeamento, Memória de Trabalho e Resolução de Problemas	55 Itens	6 aos 17 anos	Pais e professores
<i>QEF (Questionnaire of Executive Functioning)</i>	Projetado para explorar o conhecimento das crianças sobre o seu funcionamento executivo	Atenção/Concentração, Memória de Trabalho, Automonitoramento/Inibição, Mudança, Planeamento/Iniciação e Regulação Emocional	38 Itens	7 aos 14 anos	Pais e auto-relato
<i>REEF (Ratings of Everyday Executive Functioning)</i>	Serve para capturar os comportamentos de pré-escolares em contextos do quotidiano	Controlo Inibitório, Memória de Trabalho, Flexibilidade Cognitiva, Regulação Emocional e Planeamento	76 Itens	3 aos 5 anos	Pais
<i>TSI (Thinking Skills Inventory)</i>	Pretende avaliar défices de habilidades	Atenção, Memória de Trabalho, Linguagem, Comunicação, Regulação Emocional, Flexibilidade Cognitiva e Habilidades de Pensamento Social	27 Itens	6 aos 18 anos	Pais/cuidadores

3.3. Propriedades de Medida

Em relação às propriedades de medida dos instrumentos, foram avaliadas quatro características psicométricas: a Validade (i.e., grau em que um instrumento mede os construtos que pretende medir), a Confiabilidade (i.e., proporção de variância total nas medidas que se deve a diferenças entre os pacientes), a Consistência Interna (i.e., o grau de inter-relação entre os itens) e a Capacidade de Resposta (i.e., capacidade de um instrumento para detetar mudanças ao longo do tempo do construto a ser medido). Dos 41 artigos analisados, 30 analisaram a Validade, 27 avaliaram a Confiabilidade, 36 a Consistência Interna e, por fim, 6 avaliaram a Capacidade de Resposta, sendo assim a propriedade de medida menos avaliada nos estudos. Assim, os resultados mostraram também que poucos foram os estudos que testaram claramente todas as propriedades de medição, sendo que apenas 2 testaram todas as propriedades.

A classificação das características psicométricas de cada instrumento está representada na Tabela 4. Para obter esta classificação, recorreu-se ao Índice de Validade de Conteúdo (IVC) para verificar a Validade, *Alfa de Cronbach* para a Consistência Interna, ao *Kappa de Cohen* para a Confiabilidade e Coeficiente de Correlação Interclasse (CCI) para a Capacidade de Resposta, sendo usado o método de classificação dessas mesmas medidas (Cf. Anexo 2). Destarte, as características psicométricas dos instrumentos foram classificadas numa escala de cinco pontos, incluindo “muito boa”, “boa”, “razoável”, “moderada” ou “fraca”.

Tabela 3*Risco de Viés*

Instrumentos	Validade	Confiabilidade	Consistência Interna	Responsividade
BADS-C (FY Siu & Zhou)	.28	0,91 a 1,00	$\alpha = 0,94$	NA
BADS-C (Roy et. al.)	.04 a .28	K = 0.62 a 0.98	0.89	NA
BASC-2 (Dowdy et. al.)	NA	0.94 a 0.95	$\alpha = .70$	NA
BASC-2 (Karr et. al.)	NA	NA	$\alpha = .96$	NA
BDEFS-CA (Mashhadi et. al.)	.76	0.98	$\alpha = .91$	NA
BRIEF (Esbensen et. al.)	NA	NA	$\alpha = .76$ a .97	CCI = 0.50 a 0.75
BRIEF (Usher et. al.)	NA	0.79 a 0.88	0.80 a 0.98	NA
BRIEF* (Carim et. al.)	0.25	$\alpha = 0.67$ a 0.95	$\alpha = 0.90$ a 0.95	NA
BRIEF (Salas et. al.)	.62	K = .13	$\alpha = .85$ a .95	NA
BRIEF (Halvorsen et. al.)	0.93 a 0.95	K = 0.86 a 0.91	NA	NA
BRIEF (Familiar et. al.)	0.40	$\alpha = 0.90$	$\alpha = 0.93$	NA
BRIEF (Schneider et. al.)	NA	$\alpha = 0.80$ a 0.90	$\alpha = 0.80$ a 0.90	NA
BRIEF (Peters et. al.)	.62 a .67	NA	NA	CCI = 0.16 a 0.32
BRIEF	.87 a .90	0.93	$\alpha = .97$	NA

(McCoy et. al.)				
BRIEF (Reddy et. al.)	.63	0.76 a 0.85	0.80 a 0.98	NA
BRIEF (Fournet et. al)	NA	NA	$\alpha = .73$ a .92	NA
BRIEF (Fernández et. al.)	.73	NA	$\alpha = 0.92$	NA
BRIEF (Spiegel et. al.)	NA	0.80 a 0.90	0.80 a 0.97	NA
BRIEF (Huizinga & Smidts)	0.30	0.81	$\alpha = 0.78$ a 0.90	NA
BRIEF - 2 (Jiménez & Lucas-Molina)	0.88 a 0.93	$\alpha = 0.71$ a 0.91	.66 a .87	NA
BRIEF-2 (Shum et. al.)	0.26 a 0.97	NA	$\alpha = .81$ a 0.94	NA
BRIEF-2 (Jacobson et. al.)	NA	0.79	$\alpha = 0.80$ a 0.91	NA
BRIEF-P (Ezpeleta et. al.)	0.83	$\alpha = 0.86$ a 0.95	.87	0.70
BRIEF-P (Duku et. al.)	$r = .15$	0.65 a 0.94	0.80 a 0.97	NA
BRIEF-P (Neale et. al.)	.17 a .18	$K = 0.91$	NA	NA
BRIEF-P (Bonillo et. al.)	NA	NA	$\alpha = 0.9$	NA
BRIEF-SR (Byerley & Donders)	.59 a .89	$K = 0.84$	.73 a .96	NA
CEFI* (Krieger et. al.)	NA	$\alpha = 0.99$	$\alpha = 0.78$ a 0.90	CCI = .25 a .49

CBQ (Osa et. al.)	.83 a .93	NA	$\alpha = .65$	NA
CHECK (Rosenblum et. al.)	-.23 a -.62	$\alpha = .70$	$\alpha = .94$	NA
CHEXI (Camerota et. al.)	0.27 a 0.36	0.74	$\alpha = 0.85$	NA
CHEXI (Catale et. al.)	NA	.74 a .87	$\alpha = .85$ a .89	NA
CHEXI (Gutierrez et. al.)	0.65 a 0.69	0.77	$\alpha = .85$ a .89	NA
D-REF* (Warnick et. al.)	0.10 a 0.60	0.62 a 0.90	$\alpha = 0.64$ a 0.99	NA
EFORTS (Frisch & Rosenblum)	.51 a .65	NA	$\alpha = .94$	NA
EFS (Garcia et. al.)	.41 a .68	NA	$\alpha = .94$	NA
EFS (Korzeniowski & Ison)	.66	.94	$\alpha = .95$	0.1 a 0.9
LEAF (Castellanos et. al.)	0.30 a 0.49	NA	$\alpha = 0.79$	NA
QEF (Guerten et. al.)	.22 a .28	.79 a .93	.78 a .93	NA
REEF (Nilsen et. al.)	.43	NA	$\alpha = .97$	NA

TSI (Wang et. al.)	0.49 a 0.70	NA	$\alpha = 0.84$ a 0.91	NA
-----------------------	-------------	----	--------------------------	----

Notas. *Artigos adicionados posteriormente. NA = Propriedade Psicométrica não avaliada no estudo. K = *Kappa de Cohen*. α = *Alfa de Cronbach*. CCI = Coeficiente de Correlação Interclasse.

De uma forma geral, as características psicométricas apresentaram uma boa classificação, dado que a maioria foram classificadas como “boa” ou “muito boa”, significando assim que os instrumentos são válidos e confiáveis para utilização na prática clínica com crianças. Contudo, a Consistência Interna foi a que obteve melhor pontuação, pontuando, quase sempre, “muito boa”. Por outro lado, a Validade é a característica psicométrica com pior pontuação, pontuando, frequentemente, como “fraca”. No entanto, nenhum estudo apresenta boas classificações em todas as propriedades, mas também nenhum estudo apresenta piores classificações em todas as propriedades de medida (Cf. Tabela 4).

Do mesmo modo, foi avaliada a qualidade dos estudos para cada uma das propriedades psicométricas de acordo com o *Cosmin Checklist*. Esta lista de verificação contém nove propriedades de medida, no entanto, apenas foram analisadas e classificadas, as quatro principais: Validade, Confiabilidade, Consistência Interna e Capacidade de Resposta. Logo, foram apenas analisadas e respondidas, as questões relativas a estes domínios, ou seja, a *box 2* (Validade de Conteúdo), *box 3* (Validade Estrutural), *box 4* (Consistência Interna), *box 6* (Confiabilidade), *box 8* (Validade de Critério) e *box 10* (Capacidade de Resposta) (Cf. Anexo 2).

Relativamente ao modo de classificação, o grupo *Cosmin* desenvolveu uma nova escala para avaliar cada propriedade de medida. Este novo sistema de classificação é de quatro pontos, e classifica-se da seguinte forma: “Muito Bom”, “Adequado”, “Duvidoso” e “Inadequado”. Este sistema forma assim a lista de verificação *Cosmin Risk of Bias*.

Assim, no que diz respeito à qualidade dos estudos, estes obtiveram, na sua maioria, uma classificação não tão favorável, uma vez que a maioria foi classificado como “Inadequado”. Isto deve-se ao método de classificação do *Cosmin*, pois exige que se considere sempre a pior classificação (Cf. Tabela 4).

A pior classificação foi obtida, maioritariamente, na Validade, pela falta de informação relativa aos profissionais, moderadores/entrevistadores, por não referir se foi usado um guião para a entrevista, bem como a sua gravação e transcrição. Na Consistência Interna, a pior pontuação foi obtida, geralmente, por não haver o *Alfa de Cronbach*, bem como o Coeficiente de Correlação Interclasse calculados. Por sua vez, a Confiabilidade apresentou pior pontuação quando os estudos não tinham o Coeficiente de Correlação Interclasse ou correlações de *Pearson*, ou *Kappa de Cohen* calculados. Por fim, a

Capacidade de Resposta apresentava pior pontuação quando não apresentava a sensibilidade/especificidade calculadas e por não fornecer uma descrição adequada da intervenção.

Tabela 4*Classificação do Risco de Viés*

Estudos	Propriedades de Medida	Classificação	Classificação de acordo com o <i>Cosmin Checklist</i>^a
Bonillo et. al. (2012) – BRIEF-P	Validade	NA	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Boa	Duvidoso
	Capacidade de Resposta	NA	Inadequado
Byerley & Donders (2013) – BRIEF-SR	Validade	Boa	Inadequado
	Confiabilidade	Muito Boa	Duvidoso
	Consistência Interna	Boa	Duvidoso
	Capacidade de Resposta	NA	Inadequado
Camerota et. al. (2016) – CHEXI	Validade	Fraca	Adequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Muito Bom
	Capacidade de Resposta	NA	Duvidoso
Carim et. al. (2012) – BRIEF	Validade	Fraca	Inadequado
	Confiabilidade	Razoável	Inadequado
	Consistência Interna	Muito Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
Castellanos et. al. (2016) – LEAF	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Razoável	Inadequado
	Capacidade de Resposta	NA	Inadequado
Catale et. al. (2013) – CHEXI	Validade	NA	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Duvidoso
	Capacidade de Resposta	NA	Inadequado
Dowdy et. al. (2011) – BASC-2	Validade	NA	Inadequado
	Confiabilidade	Muito Boa	Inadequado
	Consistência Interna	Razoável	Muito Bom
	Capacidade de Resposta	NA	Inadequado
Duku et. al. (2014) – BRIEF-P	Validade	Fraca	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado

Esbensen et. al. (2019) – BRIEF	Validade	NA	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Razoável	Duvidoso
	Capacidade de Resposta	Moderada	Inadequado
Ezpeleta et. al. (2015) – BRIEF-P	Validade	Boa	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	Moderada	Inadequado
Familiar et. al. (2015) – BRIEF	Validade	Fraca	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Fernández et. al. (2014) – BRIEF	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Muito Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Fournet et. al. (2014) – BRIEF	Validade	NA	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
Frisch & Rosenblum (2014) – EFORTS	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Muito Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
FY Siu & Zhou (2014) – BADS-C	Validade	Fraca	Inadequado
	Confiabilidade	Muito Boa	Inadequado
	Consistência Interna	Muito Boa	Duvidoso
	Capacidade de Resposta	NA	Duvidoso
Garcia et. al. (2016) – EFS	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Muito Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Guerten et. al. (2016) – QEF	Validade	Fraca	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Gutierrez et. al. (2021) – CHEXI	Validade	Fraca	Duvidoso
	Confiabilidade	Boa	Inadequado

Halvorsen et. al. (2018) – BRIEF	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Duvidoso
	Validade	Muito Boa	Inadequado
	Confiabilidade	Muito Boa	Inadequado
Huizinga & Smidts (2011) – BRIEF	Consistência Interna	NA	Inadequado
	Capacidade de Resposta	NA	Inadequado
	Validade	Fraca	Inadequado
	Confiabilidade	Muito Boa	Inadequado
Jacobson et. al. (2019) – BRIEF-2	Consistência Interna	Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
	Validade	NA	Inadequado
	Confiabilidade	Boa	Inadequado
Jiménez & Lucas- Molina (2019) – BRIEF-2	Consistência Interna	Boa	Duvidoso
	Capacidade de Resposta	NA	Inadequado
	Validade	Boa	Inadequado
	Confiabilidade	Boa	Inadequado
Karr & Garcia-Barrera (2016) – BASC-2	Consistência Interna	Razoável	Muito Bom
	Capacidade de Resposta	NA	Inadequado
	Validade	NA	Inadequado
	Confiabilidade	NA	Inadequado
Korzeniowski & Ison (2019) – EFS	Consistência Interna	Muito Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
	Validade	Fraca	Inadequado
	Confiabilidade	Muito Boa	Duvidoso
Krieger et. al. (2019) – CEFI	Consistência Interna	Muito Boa	Muito Bom
	Capacidade de Resposta	Fraca	Inadequado
	Validade	NA	Inadequado
	Confiabilidade	Muito Boa	Inadequado
Mashhadi et. al. (2020) – BDEFS-CA	Consistência Interna	Boa	Muito Bom
	Capacidade de Resposta	Fraca	Inadequado
	Validade	Fraca	Duvidoso
	Confiabilidade	Muito Boa	Inadequado
McCoy et. al. (2011) – BRIEF	Consistência Interna	Muito Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
	Validade	Boa	Inadequado
	Confiabilidade	Muito Boa	Inadequado
	Consistência Interna	Muito Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado

Neale et. al. (2018) – BRIEF-P	Validade	Fraca	Inadequado
	Confiabilidade	Muito Boa	Inadequado
	Consistência Interna	NA	Inadequado
	Capacidade de Resposta	NA	Inadequado
Nilsen et. al. (2016) – REEF	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Muito Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Osa et. al. (2013) – CBQ	Validade	Boa	Inadequado
	Confiabilidade	NA	Duvidoso
	Consistência Interna	Fraca	Muito Bom
	Capacidade de Resposta	NA	Inadequado
Peters et. al. (2012) – BRIEF	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	NA	Inadequado
	Capacidade de Resposta	Fraca	Inadequado
Reddy et. al. (2011) – BRIEF	Validade	Fraca	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	NA	Duvidoso
	Capacidade de Resposta	Boa	Duvidoso
Rosenblum et. al. (2019) – CHECK	Validade	Fraca	Inadequado
	Confiabilidade	Razoável	Inadequado
	Consistência Interna	Muito Boa	Duvidoso
	Capacidade de Resposta	NA	Duvidoso
Roy et. al. (2015) – BADSC	Validade	Fraca	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Duvidoso
	Capacidade de Resposta	NA	Inadequado
Salas et. al. (2016) – BRIEF	Validade	Fraca	Inadequado
	Confiabilidade	Fraca	Inadequado
	Consistência Interna	Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
Schneider et. al. (2019) – BRIEF	Validade	NA	Inadequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado

Shum et. al. (2020) – BRIEF-2	Validade	Boa	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Spiegel et. al. (2020) – BRIEF-2	Validade	NA	Inadequado
	Confiabilidade	Muito Boa	Inadequado
	Consistência Interna	Boa	Inadequado
	Capacidade de Resposta	NA	Inadequado
Usher et. al. (2015) – BRIEF	Validade	NA	Adequado
	Confiabilidade	Boa	Inadequado
	Consistência Interna	Boa	Duvidoso
	Capacidade de Resposta	NA	Duvidoso
Wang et. al. (2018) – TSI	Validade	Fraca	Inadequado
	Confiabilidade	NA	Inadequado
	Consistência Interna	Boa	Muito Bom
	Capacidade de Resposta	NA	Inadequado
Warnick et. al. (2015) – D-REF	Validade	Fraca	Inadequado
	Confiabilidade	Razoável	Inadequado
	Consistência Interna	Razoável	Duvidoso
	Capacidade de Resposta	NA	Inadequado

Nota. NA= Propriedade psicométrica não avaliada no estudo. *A classificação final do estudo para cada propriedade psicométrica, foi obtida considerando-se a pior classificação alcançada para um dos itens específicos de cada propriedade.

4. Discussão

O presente estudo objetivou, a partir de uma revisão sistemática da literatura, verificar quais os instrumentos de avaliação das FE existem para crianças, nomeadamente questionários e inventários. Objetivou-se também a verificação das características psicométricas dos instrumentos e a qualidade dos seus estudos. O seu propósito é identificar e explorar a literatura existente, de forma a ser possível identificar eventuais lacunas e fornecer novas perspetivas para a investigação futura. Justifica-se o presente estudo na medida em que não existem estudos que confirmam e/ou compilam quais os questionários/inventários existem para a população infantil, pelo que a maioria dos estudos se refere a outros instrumentos de avaliação das FE, tais como testes e baterias.

4.1. Discussão dos principais resultados

4.1.1. Características dos estudos/instrumentos

Perante os resultados deste estudo foi possível dar conta de um único artigo de validação do questionário BRIEF para português do Brasil, não sendo, dessa forma, apresentado nenhum validado para o português de Portugal. De acordo com a literatura há uma lacuna no que toca a instrumentos de avaliação das FE em crianças, em língua portuguesa de Portugal (Santana et al. 2019).

Contudo, a literatura demonstra que, apesar de o uso dos questionários se limitar principalmente aos países de língua inglesa, já começam a surgir estudos de tradução destes instrumentos para o português (Carim et al., 2012). Até agora parece existir duas exceções, sendo o BRIEF e o CHEXI, os instrumentos para os quais foram realizados estudos de validação usando versões em vários idiomas diferentes, nomeadamente para o português (Catale et al., 2013). Contudo, não existe muita informação disponível, nem muitos estudos. A título de exemplo, o CHEXI, só é possível ter acesso ao inventário, não existindo quaisquer estudos disponíveis sobre a tradução (Moura et al., 2008). Dessa forma, não há informação, nem nenhuma descrição de quais critérios foram usados para a realização da respetiva tradução.

Para além de haver poucos estudos traduzidos para o português, também se verifica que muitos questionários de avaliação das FE são recentes, pelo que os estudos foram publicados nos últimos anos (Karr & Garcia-Barrera, 2016).

Os questionários são possíveis de avaliar nas mais diversas idades, nomeadamente, desde os 2-21 anos, pelo que é um ponto positivo. Contudo, há poucos questionários dirigidos a crianças pequenas, mais precisamente, dos 2-5 anos, algo preocupante, uma vez que as FE surgem já no início do primeiro ano de vida, desenvolvendo-se, posteriormente, por volta dessa idade (Zelazo & Muller, 2002). A literatura confirma que ainda é incomum avaliar crianças menores de 6 anos de idade, e isto deve-se, em parte, à escassez de ferramentas adequadas para avaliar as FE em crianças pequenas (O'Mehager et al., 2019). Contudo, os anos pré-escolares foram identificados como um momento particularmente crucial, tanto no surgimento como no desenvolvimento das FE (Howard et al., 2015). Por isso, melhorar e avaliar as FE no início da vida é importante, uma vez que estas são fundamentais, posteriormente, principalmente para o sucesso escolar (Diamond, 2012).

Para além disto, é notório que o BRIEF é o questionário mais usado para avaliar as FE em crianças, pois mostra ser uma escala de fácil aplicação, com confiabilidade adequada, sugerindo ser um instrumento importante de avaliação. É um instrumento com 86 itens que serve para avaliar crianças dos 5 aos 18 anos, sendo que ainda tem uma segunda edição e uma versão para o pré-escolar dos 2 aos 5 anos, com 63 itens (Huizinga & Smidts, 2011). Isto confirma o que a literatura relata, o BRIEF é o questionário mais comumente usado para avaliar as FE em crianças (Schneider et al., 2019).

4.1.2. Propriedades de medida

Relativamente às propriedades de medida, a Validade, a Confiabilidade e a Consistência Interna foram as mais investigadas, sendo que poucos estudos avaliaram a Capacidade de Resposta dos instrumentos. Algo semelhante ao que ocorreu no estudo de Campos et al. (2018), uma vez que nenhum dos instrumentos usados nesse estudo tinham sido avaliados quanto a esta propriedade de medida. Contudo, esta também é necessária ser avaliada, pois trata-se de um elemento crucial para avaliar a validade do instrumento, mas também porque consiste em testar a capacidade de um instrumento de medir mudanças clinicamente importantes (Couto et al., 2019).

Para Souza et al. (2017), a Validade e a Confiabilidade, não são totalmente independentes uma da outra, no entanto, existe um consenso unânime entre os pesquisadores em considerar estas duas como principais propriedades de medida de instrumentos. Isto porque o bom desempenho dos resultados dos instrumentos se deve, em parte, a estas medidas. Contudo, e segundo o mesmo autor, a Consistência Interna, é igualmente uma importante propriedade de medida de instrumentos (Souza et al., 2017).

Em relação à aplicabilidade dos instrumentos, deve-se ter em atenção na sua aplicação pelo que não têm as suas propriedades de medida totalmente testadas (Garcia et al., 2016). Estes instrumentos de medida integram a prática clínica e, por isso, a avaliação da sua qualidade é fundamental para a seleção de instrumentos que forneçam medidas válidas e confiáveis (Souza et al., 2017).

4.1.3. Ferramenta *Cosmin Checklist*

Para a avaliação da qualidade dos estudos, foi utilizada a lista de verificação *Cosmin Checklist* pelo facto de que os seus critérios de qualidade são os mais atualizados e amplamente aceites na literatura. O número de revisões sistemáticas de instrumentos de avaliação como os questionários têm vindo a aumentar, e a lista de verificação *Cosmin* é frequentemente usada em muitas dessas revisões (Mokkink et al., 2018).

No entanto, muitos pesquisadores sugeriram várias melhorias na lista de verificação, sendo que a última foi realizada em 2018 (Mokkink et al., 2018). Segundo Mokking et al. (2018), alguns estudos chegaram a ser classificados como “inadequados” por terem um tamanho de amostra pequeno, por exemplo. Situação essa que foi possível verificar neste estudo, uma vez que a maioria dos estudos obtiveram uma fraca classificação. Caso o estudo não apresentasse informação relativa à existência de guião de entrevista, por exemplo, já seria motivo para obter uma má classificação nesse domínio. Tal como Campos et al. (2018) refere, mesmo que algum item de avaliação de determinada propriedade seja “excelente”, se houver outro que seja “fraco” a classificação final para aquela propriedade será “fraco” (Campos et al., 2018).

4.2. Limitações e Implicações Clínicas

Este estudo possui algumas limitações, tais como o reduzido tempo para análise dos estudos e dos questionários e o facto de a maioria dos estudos não avaliarem todas as propriedades de medida dos seus instrumentos, o que dificultou a análise e avaliação dos mesmos. A par disto, é identificada também como uma limitação o facto de não ter sido avaliadas todas as propriedades de medida, neste estudo, o que permitiria uma melhor análise dos instrumentos e uma conclusão mais fiável. Também a abundante frequência de estudos sobre o BRIEF e o pouco estudo de outros questionários, dificultou a perceção da totalidade dos instrumentos existentes para avaliar as FE em crianças e o objetivo final deste estudo.

Por outro lado, também é possível identificar limitações da ferramenta de avaliação da qualidade dos estudos, o *Cosmin Checklist*, pelo que exige que se tenha sempre em conta a pior classificação alcançada para um dos itens específicos de cada propriedade, o que leva a que a maioria obtenha uma má classificação, não permitindo uma melhor leitura acerca da qualidade dos estudos/instrumentos.

Como implicações clínicas, os achados desta revisão apontam para os instrumentos disponíveis na literatura com boas características psicométricas, estando assim disponíveis para a mensuração das FE em crianças. No entanto, os resultados da qualidade dos estudos não são os mais favoráveis. Portanto, embora estes instrumentos se encontrem disponíveis, os profissionais deverão selecionar o instrumento mais adequado tendo em conta as suas características psicométricas.

Para estudos futuros, e tendo em conta as limitações mencionadas, recomenda-se que os pesquisadores considerem a realização da avaliação completa das propriedades de medida dos instrumentos, de forma a torná-los ainda mais robustos e confiáveis. Recomenda-se também que sejam realizados mais estudos dos restantes instrumentos de avaliação às quais existem poucos estudos, tais como o EFORTS, TSI e o CHECK, de forma a haver uma melhor compreensão acerca destes instrumentos, que são igualmente importantes para avaliar as FE em crianças. Por fim, recomenda-se a tradução e validação destes instrumentos para a língua portuguesa, pois será um grande contributo para a prática clínica.

5. Conclusão

A investigação apresentada destinou-se a aprofundar o tema da avaliação das FE em crianças por meio de questionários/inventários e a avaliação das suas características psicométricas. Em jeito de conclusão, este estudo retirou as seguintes conclusões:

Primeiramente, esta revisão sistemática da literatura confirma a existência de uma grande diversidade de questionários/inventários que avaliam as FE em crianças. No entanto, considera-se que, apesar desta diversidade, esta área ainda conta com pouca investigação, devido à recente literatura e por se prender bastante a um único instrumento (i.e. BRIEF).

Além disso, este estudo testemunha a escassa quantidade de questionários/inventários traduzidos para a língua portuguesa e constata a importância dessas traduções para a prática clínica. No entanto, ainda demonstra a importância de analisar e avaliar as características psicométricas dos instrumentos, bem como a qualidade dos seus estudos.

Em suma, os resultados da presente revisão demonstraram que o uso de questionários traz uma importante contribuição para o campo de pesquisa, pois foi demonstrado que abordam pelo menos parcialmente diferentes aspetos das FE (Catale et al., 2013).

Referências Bibliográficas

- *Abreu, P., Nikaedo, C., Abreu, N., Tourinho, C., Miranda, M., Bueno, O., & Martin, R. (2014). Working memory screening, school context, and socioeconomic status: an analysis of the effectiveness of the working memory rating scale in brazil. *Journal of Attention Disorders*, 18(4), 346-356. <http://dx.doi.org/10.1177/1087054713476138>
- Anderson, V., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of executive functions through late childhood and adolescence in an australian sample. *Development Neuropsychology*, 20(1), 385-406. http://dx.doi.org/10.1207/S15326942DN2001_5
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8(2), 71-82. <https://doi.org/10.1076/chin.8.2.71.8724>
- *Annotti, L., & Teglasi, H. (2016). Functioning in the real world: using storytelling to improve validity in the assessment of executive functions. *Journal of Personality Assessment*, 1-12. <http://dx.doi.org/10.1080/00223891.2016.1205075>
- Baddeley, A. D., & Hitch, G. J. (1974). *Working memory*. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–87). London: Academic Press [http://dx.doi.org/10.1016/s0079-7421\(08\)60452-1](http://dx.doi.org/10.1016/s0079-7421(08)60452-1)
- *Bakar, E., Taner, Y., Soysal, A., Karakas, S., & Turgay, A. (2011). Behavioral rating inventory and laboratory tests measure different aspects of executive functioning in boys: a validity study. *Klinik Psikofarmakoloji Bülteni- Bulletin of Clinical Psychopharmacology*, 21(4), 302-316. <https://doi.org/10.5455/bcp.20111004014003>
- *Bardack, S., & Obradovic, J. (2019). Observing teachers' displays and scaffolding of executive functioning in the classroom context. *Journal of Applied Developmental Psychology*, 62, 205-219. <https://doi.org/10.1016/j.appdev.2018.12.004>
- *Bat, M., Ishai, R., & Levi, N. (2014). Validating PPAT's symbolic meanings of emotional and cognitive functioning among children. *The Arts in Psychotherapy*, 41(4), 309-319. <http://dx.doi.org/doi:10.1016/j.aip.2014.05.001>

- *Bat, M. (2014). FEATS Problem-solving scale in PPAT of children aged 5–6.5 as related to their executive functions and motivation. *The Arts in Psychotherapy*, 41, 27-35. <http://dx.doi.org/10.1016/j.aip.2013.11.008>
- *Berg, L., Weerd, A., Reuvekamp, H., & Meere, J. (2019). The burden of parenting children with frontal lobe epilepsy. *Epilepsy & Behavior*, 97, 269-274. <https://doi.org/10.1016/j.yebeh.2019.05.034>
- *Bemes, G., Villodas, M., Coles, C., Kable, J., May, P., Kalberg, W., Sowell, E., Jones, K., Riley, E., & Mattson, S. (2021). Validity and reliability of executive function measures in children with heavy prenatal alcohol exposure: correspondence between multiple raters and laboratory measures. *Alcoholism: Clinical and Experimental Research*, 45(3), 596-607. <http://doi.org/10.1111/acer.14547>
- *Blijd-Hoogewys, E., Bezemer, M., & Geert, P. (2014). Executive functioning in children with ASD: an analysis of the BRIEF. *Journal of Autism and Developmental Disorders*, 44, 3089-3100. <http://doi.org/10.1007/s10803-014-2176-9>
- *Bonillo, A., Jiménez, E., Ballabriga, M., Capdevila, C., & Riera, R. (2012). Validation of catalan version of BRIEF-P. *Child Neuropsychology*, 18(4), 347-355. <http://dx.doi.org/10.1080/09297049.2011.613808>
- *Byerley, A., & Donders, J. (2013). Clinical utility of the behavior rating inventory of executive function-self-report (BRIEF-SR) in adolescents with traumatic brain injury. *Rehabilitation Psychology*, 58(4), 412-421. <http://dx.doi.org/10.1037/a0034228>
- *Camerota, M., T. Willoughby, M., J.Kuhm, L., & B. Blair, C. (2016). The childhood executive functioning inventory (CHEXI): factor structure, measurement invariance, and correlates in US preschoolers. *Child Neuropsychology*, 1-16. <http://dx.doi.org/10.1080/09297049.2016.1247795>
- Campos, M., Felicidade, C., Aguiar, S., Vieira, D. (2018). Psychometric properties of physical activity questionnaires in adolescence: systematic review. *Revista Brasileira de Atividade Física & Saúde*, 1-13. <http://dx.doi.org/10.12820/rbafs.23e0045>

- *Carim, D., C. Miranda, M., & Bueno, O. (2012). Tradução e adaptação para o português do “*Behavior Rating Inventory of Executive Function*”. *Psicologia: Reflexão e Crítica*, 25(4), 653-661. <https://doi.org/10.1590/S0102-79722012000400004>
- *Castellanos, I., Kronenberger, W., & Pisoni, D. (2016). Questionnaire-based assessment of executive functioning: psychometrics. *Applied Neuropsychology: Child*, 1-18. <http://dx.doi.org/10.1080/21622965.2016.1248557>
- *Catale, C., Meullemans, T., & B. Thorell, L. (2013). The childhood executive function inventory: confirmatory factor analyses and cross-cultural clinical validity in a sample of 8- to 11-year-old children. *Journal of Attention Disorders*, 1-7. <http://dx.doi.org/10.1177/1087054712470971>
- *Chevignard, M., Kerrouche, B., Krasrey-Pacini, A., Mariller, A., Pineau-Chardon, E., Notteghem, P., Fournet, N., & Roy, A. (2017). Ecological assessment of everyday executive functioning at home and at school following childhood traumatic brain injury using the BRIEF questionnaire. *Journal of Head Trauma Rehabilitation*, 1-12. <http://dx.doi.org/10.1097/HTR.0000000000000295>
- *Chubarov, E., Sommerfeld, E., Hermesh, H., Shoval, G., Weiser, M., & Zalsman, (2015). Dynamic occupation assessment of executive function in adolescents with schizophrenia spectrum disorders: an initial report. *European Psychiatry*, 30, 894-899. <http://dx.doi.org/10.1016/j.eurpsy.2015.09.001>
- Couto, R., Leal, F., Pitta, G., & Andreoni, S. (2019). Responsiveness of the CCVUQ-Br quality of life questionnaire in chronic venous ulcer patients. *Jornal Vascular Brasileiro*, 1-9. <https://doi.org/10.1590/1677-5449.190047>
- *d’Ardhuy, X., Edgin, J., Bouis, C., Sola, S., Goeldner, C., Kishnani, P., Noldeke, J., Rice, S., Sacco, S., Squassante, L., Spiridigliozzi, G., Visootsak, J., Heller, J., & Khwaja, O. (2015). Assessment of cognitive scales to examine memory, executive function and language in individuals with down syndrome: implications of a 6-month observational study. *Frontiers in Behavioral Neuroscience*, 9, 1-11. <http://dx.doi.org/10.3389/fnbeh.2015.00300>

- *Davidson, F., Cherry, K., & Corkum, P. (2015). Validating the behavior rating inventory of executive functioning for children with ADHD and their typically developing peers. *Applied Neuropsychology: Child*, 1-11. <http://dx.doi.org/10.1080/21622965.2015.1021957>
- *Dehn, M. (2014). The children's psychological processes scale: factor structure and correlations with performance-based measures. *Journal of Attention Disorders*, 18(4), 305-317. <http://dx.doi.org/10.1177/1087054714521294>
- *Dekker, M., Ziermans, T., Spruijt, A., & Swaab, H. (2017). Cognitive, parent and teacher rating measures of executive functioning: shared and unique influences on school achievement. *Frontiers in Psychology*, 8, 1-13. <http://dx.doi.org/10.3389/fpsyg.2017.00048>
- Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, 21(5), 335-341. <http://dx.doi.org/10.1177/0963721412453722>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135-168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dias, N., Menezes, A., & Seabra, A. (2010). Alterações das funções executivas em crianças e adolescentes. *Estudos Interdisciplinares em Psicologia*, 1(1), 80-95. <http://dx.doi.org/10.5433/2236-6407.2010v1n1p80>
- *Dowdy, E., K. Chin, J., M. Twyford, J., V. Dever, B. (2011). A factor analytic investigation of the BASC-2 behavioral and emotional screening system parent form: psychometric properties, practical implications and future directions. *Journal of School Psychology*, 49, 265-280. <http://dx.doi.org/10.1016/j.jsp.2011.03.005>
- *Duku, E., & Vaillancourt, T. (2014). Validation of the BRIEF-P in a sample of a canadian preschool children. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 20(3), 358-371. <http://dx.doi.org/10.1080/09297049.2013.796919>
- *Esbensen, A., Hoffman, E., Shaffer, R., Chen, E., Patel, L., & Jacola, L. (2019). Reliability of informant-report measures of executive functioning in children with down

- syndrome. *American Journal on Intellectual and Development Disabilities*, 124(3), 220-233. <http://dx.doi.org/10.1352/1944-7558-124.3.220>
- *Ezpeleta, L., Granero, R., Penelo, E., Osa, N., & M. Domènech, J. (2015). Behavior rating inventory of executive functioning-preschool (BRIEF-P) applied to teachers: psychometric properties and usefulness for disruptive disorders in 3-year-old preschoolers. *Journal of Attention Disorders*, 19(6), 476-488. <http://dx.doi.org/0.1177/1087054712466439>
- *Familiar, I., Ruisenor-Escudero, H., Giordani, B., Bangirana, P., Nakasujja, N., Opoka, R., & Boivin, M. (2015). Use of the behavior rating inventory of executive function and child behavior checklist in ugandan children with HIV or a history of severe malaria. *Journal of Development and Behavioral Pediatrics*, 36(4), 277-284. <http://dx.doi.org/10.1097/DBP.0000000000000149>
- *Faridi, N., Karama, S., Burgaleta, M., White, M., C.Evans, A., & Waber, D. (2015). Neuroanatomical correlates of behavioral rating versus performance measures of working memory in typically developing children and adolescents. *Neuropsychology*, 29(1), 82-91. <http://dx.doi.org/10.1037/neu0000079>
- *Fernández, T., González-Pienda, J., Pérez, C., García, D., & Pérez, L. (2014). Psychometric characteristics of the BRIEF scale for the assessment of eecutive functions in spanish clinical population. *Psicohtema*, 26(1), 47-52. <http://dx.doi.org/0.7334/psicothema2013.149>
- *Fournet, N., Roulin, J., Monnier, C., Atzeni, T., Cosnefroy, O., Gall, D., & Roy, A. (2014). Multigroup confirmatory factor analysis ans structural invariance with age of the behavior rating inventory of executive function (BRIEF) – french version. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 1-22. <http://dx.doi.org/10.1080/09297049.2014.906569>
- *Frisch, C., & Rosenblum, S. (2014). Reliability and validity of the executive function and occupational routines scales (EFORTS). *Research in Development Disabilities*, 35, 2148-2157. <http://dx.doi.org/10.1016/j.ridd.2014.05.003>

- *F.Y, Siu, A., & Zhou, Y. (2014). Behavioral assessment of the dysexecutive syndrome for children: an examination of clinical utility for children with attention- deficit hyperactivity disorder (ADHD). *Journal of Child Neurology*, 29(5), 608-616. <http://doi.org/10.1177/0883073813516191>
- *Gabrielli, J., Jackson, Y., & Brown, S. (2014). Measurement of behavioral and emotional outcomes of youth in foster care: investigation of the roles of age and placement type. *Journal of Psychopathology and Behavioral Assessment*, 1-10. <http://dx.doi.org/10.1007/s10862-014-9464-8>
- *García, T., Álvarez-García, D., Areces, D., Seguro, L., González-Castro, P., & Rodríguez, C. (2016). Psychometric characteristics of the executive functioning scale for families (EFS-F). *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 47(2), 71-82. <https://doi.org/10.21865/RIDEP47.2.05>
- *García-Barrera, M., Karr, J., & W. Kamphaus, R. (2013). Longitudinal applications of a behavioral screener of executive functioning: assessing factorial invariance and exploring latent growth. *Psychological Assessment*, 25(4), 1300-1313. <http://doi.org/10.1037/a0034046>
- *Garcia-Barrera, M., Kamphaus, R., & Baudalos, D. (2011). Theoretical and Statistical Derivation of a Screener for the Behavioral Assessment of Executive Functions in Children. *Psychological Assessment*, 23(1), 64-79. <http://doi.org/10.1037/a0021097>
- Gazzaniga, M., Irvy, R. B., & Mangun, G. R. (1998). *Cognitive neuroscience. The biology of the mind*. New York: WW Norton.
- *Guerten, M., Catale, C., Guerten, C., Wansard, M., & Meulemans, T. (2016). Studying self-awareness in children: validation of the questionnaire of executive functioning (QEF). *The Clinical Neuropsychologist*, 1-22. <http://dx.doi.org/10.1080/13854046.2016.1178331>
- *Gilboa, Y., Jansari, A., Kerrouche, B., Uçak, E., Tiberghien, A., Benkhaled, O., Aligon, D., Mariller, A., Verdier, V., Mintegui, A., Abada, G., Canizares, C., Goldstein, A., & Chevignard, M. (2017). Assessment of executive functions in children and adolescents with acquired brain injury (ABI) using a novel complex multi-tasking

- computerised task: The Jansari assessment of Executive Functions for Children (JEF-C©). *Neuropsychological Rehabilitation*, 1-25.
<https://doi.org/10.1080/09602011.2017.1411819>
- *Gilboa, Y., Rosenblum, S., Fattal-Valevski, A., Toledano-Alhadeff, H., & Josman, N. (2014). Is there a relationship between executive functions and academic success in children with neurofibromatosis type 1? *Neuropsychological Rehabilitation: An International Journal*, 24(6), 918-935.
<http://dx.doi.org/10.1080/09602011.2014.920262>
- *Gil-Hernandez, S., & Garcia-Moreno, L. Executive performance and dysexecutive symptoms in binge drinking adolescents. *Alcohol*, 51, 79-87.
<http://dx.doi.org/10.1016/j.alcohol.2016.01.003>
- Goldstein, S., Naglieri, J. (2014). *Handbook of executive functioning*.
http://dx.doi.org/10.1007/978-1-4614-8106-5_1
- Goldstein, S., Naglieri, J.A., Princiotta, D., & Otero, T.M. (2014). Introduction: a history of executive functioning as a theoretical and clinic construct. In S. Goldstein & J.A.Naglieri (Eds.) *Handbook of executive function* (pp. 3-12). New york: Springer.
http://dx.doi.org/10.1007/978-1-4614-8106-5_1
- *Gutierrez, M., Filippetti, V., & Lemos, V. (2021). The childhood executive functioning inventory (CHEXI) parent and teacher form: factor structure and cognitive correlates in spanish – speaking children from argentina. *Development Neuropsychology*, 46(2), 136-148.
<https://doi.org/10.1080/87565641.2021.1878175>
- *Haack, L., & Araujo, E. (2017). Culturally appropriate assessment of functioning in diverse children: development and preliminary validation of the FX-II scale in mexico. *Journal of Attention Disorders*, 1-15. <http://dx.doi.org/10.1177/1087054717730613>
- Halse, M., Steinsbekk, S., Hammar, A. Belsky, J., & Wichstrom, L. (2019). Parental predictors of children's executive functioning from ages 6 to 10. *British Journal of Developmental Psychology*, 37, 410-426. <http://doi.org/10.1111/bjdp.12282>
- *Halvorsen, M., Mathiassen, B., Amundsen, T., Ellingsen, J., Brondbo, P., Sundby, J., Steinsvik, O., Martinussen, M. (2018). Confirmatory factor analysis of the behavior

- rating inventory of executive function in a neuro-pediatric sample and its application to mental disorders. *Child Neuropsychology*, 1-19. <https://doi.org/10.1080/09297049.2018.1508564>
- Howard, S., Okely, A., & Ellis, Y. (2015). Evaluation of a differentiation model of preschoolers' executive functions. *Frontiers in Psychology*, 6(285), 1-8. <http://doi.org/10.3389/fpsyg.2015.00285>
- *Huizinga, M., & Smidts, D. P. (2011). Age-related changes in executive function: A normative study with the dutch version of the behavior rating inventory of executive function (BRIEF). *Child Neuropsychology*, 17(1), 51–66. <https://doi.org/10.1080/09297049.2010.509715>
- *Jacobson, L., G. Kalb, L., & Mahone, E. (2019). When theory met data: factor structure of the BRIEF-2 in a clinical sample. *The Clinical Neuropsychologist*, 1-17. <http://dx.doi.org/10.1080/13854046.2019.1571634>
- *Jiménez, A., & Lucas-Molina, B. (2018). Dimensional structure and measurement invariance of the BRIEF-2 across gender in a socially vulnerable sample of primary school-aged children. *Child Neuropsychology*, 1-13. <https://doi.org/10.1080/09297049.2018.1512962>
- Kamradt, J., Nikolas, M., Burns, G., Garner, A., Jarrett, M., Luebbe, A., & Becker, S. (2019). Barkley deficits in executive functioning scale (BDEFS): validation in a large multisite college sample. *Assessment*, 1-13. <http://dx.doi.org/10.1177/1073191119869823>
- *Karr, J., & Garcia-Barrera, M. (2016). The assessment of executive functions using the BASC-2. *Psychological Assessment*, 1-17. <http://dx.doi.org/10.1037/pas0000418>
- *Kavanaugh, B., Cancilliere, M., Fryc, A., Tirrell, E., Oliveira, J., Oberman, L., Wexler, B., Carpenter, L., & Spirito, A. (2019). Measurement of executive functioning with the national institute of health toolbox and the association to anxiety/depressive symptomatology in childhood/adolescence. *Child Neuropsychology*, 1-17. <https://doi.org/10.1080/09297049.2019.1708295>

- *Kechter, A., Black, D., Riggs, N., Warren, C., Ritt- Olson, A., Chou, C., & Pentz, M. (2019). Factors in the perceived stress scale differentially associate with mindfulness disposition and executive function among early adolescents. *Journal of Child and Family Studies*, 1-8. <https://doi.org/10.1007/s10826-018-01313-4>
- *Keiser, A., & Reddy, L. (2013). Clinical use of the pediatric attention disorders diagnostic screener for children at risk for attention deficit hyperactivity disorder: case illustrations. *Journal of Applied School Psychology*, 29(4), 350-374. <http://dx.doi.org/10.1080/15377903.2013.836777>
- Kihlstrom, J. (2021). Ecological validity and “ecological validity”. *Perspectivas sobre a ciência psicológica*, 16(2), 466-471. <http://dx.doi.org/10.1177/745691620966791>
- *Klenberg, L., Hokkanen, L., Lahti-Nuuttila, P., & Narhi, V. (2016). Teacher ratings of executive function difficulties in finnish children with combined and predominantly inattentive symptoms of ADHD. *Applied Neuropsychology: Child*, 1-11. <http://dx.doi.org/10.1080/21622965.2016.1177531>
- *Korzeniowski, C., & Ison, M. (2019). Escala de funcionamiento ejecutivo para escolares: análisis de las propiedades psicométricas. *Psicología Educativa*, 25(2), 147-157. <https://doi.org/10.5093/psed2019a4>
- *Krieger, V., Amador-Campos, J., & Però-Cebollero, M. (2019). Interrater agreement on behavioral executive function measures in adolescents with attention deficit hyperactivity disorder. *International Journal of Clinical and Health Psychology*, 19, 141-149. <https://doi.org/10.1016/j.ijchp.2019.02.007>
- *Krivitzky, L., Bosenbark, D., Ichord, R., Jastrzab, L., & Billingshurst, L. (2019). Brief report: relationship between performance testing and parent report of attention and executive functioning profiles in children following perinatal arterial ischemic stroke. *Child Neuropsychology*, 1-10. <https://doi.org/10.1080/09297049.2019.1588957>
- Levine, B., & Craik, F. I. (2012). *Unifying clinical, experimental, and neuroimaging studies of the human frontal lobes*. In B. Levine & F. I. Craik (Eds.), *Mind and the frontal*

- lobes: Cognition, behavior, and brain imaging (pp. 3–15). New York, NY: Oxford University Press
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment* (4th ed.). New York: Oxford University Press
- *Magimairaj, B. (2018). Parent-rating vs performance-based working memory measures: association with spoken language measures in school-age children. *Journal of Communication Disorders*, 1-41. <https://doi.org/10.1016/j.jcomdis.2018.09.001>
- *Mashhadi, A., Maleki, Z., Hasani, J., & Tabatabaei, S. (2020). Psychometric properties of persian version of the barkley deficits in executive functioning scale–children and adolescents. *Applied Neuropsychology: Child*, 1-9. <https://doi.org/10.1080/21622965.2020.1726352>
- *McCoy, D., Raver, C., Lowenstein, A., Tirado-Strayer, N. (2011). Assessing self-regulation in the classroom: validation of the BIS-11 and the BRIEF in low-income, ethnic minority school-age children. *Early Education and Development*, 22(6), 883-906. <http://dx.doi.org/10.1080/10409289.2010.508371>
- McGuinness, L., & Higgins, J. (2020). Risk-of-bias Visualization (robvis): An r package and shiny web app for visualizing risk-of-bias assessments. *Research Sinthesis Methods*, 1-7. <http://doi.org/10.1002/jrsm.1411>
- Miyake, A., Friedman, N., Emerson, M., Witzki, A., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive Psychology*, 41, 49-100. <http://dx.doi.org/10.1006/cogp.1999.0734>
- Mokking, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., L. Knol, D., M. Bouter, L., & C. W. de Vet, H. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, 19, 539-549. <http://doi.org/10.1007/s11136-010-9606-8>
- Mokking, L., C.W. de Vet, H., Prinsen, C., Patrick, D., Alonso, J., M. Bouter, L., & Terwee, C. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-

- Reported Outcome Measures. *Quality of Life Research*, 27, 1171-1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Moura, O., Leite, M., Albuquerque, C., Vilar, M., Santos, M., Lopes, A., Pereira, M., & Simões, M. (2008). Inventário de funcionamento de crianças (CHEXI) para pais e professores. Disponível em: https://chexi.se/onewebmedia/CHEXI_Portuguese.pdf
- *Neale, D., Basilio, M., & Whitebread, D. (2018). The grasping task: A 12-month predictor of 24-month delay task performance and BRIEF-P inhibition scores. *Infant and Child Development*, 1-11. <https://doi.org/10.1002/icd.2092>
- Nesbitt, K. T., & Farran, D. C. (2021). Effects of prekindergarten curricula: Tools of the Mind as a case study. *Monographs of the Society for Research in Child Development*, 86(1), 1-116. <https://doi.org/10.1111/mono.12425>
- *Nilsen, E., Huyder, V., McAuley, T., & Liebermann, D. (2016). Ratings of everyday executive functioning (REEF): a parent-report measure of preschoolers' executive functioning skills. *Psychological Assessment*, 1-16. <http://dx.doi.org/10.1037/pas0000308>
- *O'Meagher, A., Norris, K., Kemp, N., & Anderson, P. (2019). Parent and teacher reporting of executive function and behavioral difficulties in preterm and term children at kindergarten. *Applied Neuropsychology: Child*, 1-13. <http://doi.org/10.1080/21622965.2018.1550404>
- *Osa, N., Granero, R., Penelo, E., Domènech, J., & Ezpeleta, L. (2013). The short and very short forms of the children's behavior questionnaire in a community sample of preschoolers. *Assessment*, 1-14. <http://dx.doi.org/10.1177/1073191113508809>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – a web and mobile app for systematic reviews. *Ouzzani et al. Systematic Reviews*, 5(210), 1-10. <http://dx.doi.org/10.1186/s13643-016-0384-4>
- Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71), 1-9. <https://doi.org/10.1136/bmj.n71>

- *Payne, J., Hyman, S., Shores, E., & North, K. (2011). Assessment of executive function and attention in children with neurofibromatosis type 1: relationships between cognitive measures and real-world behavior. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 17(4), 313-329. <http://dx.doi.org/10.1080/09297049.2010.542746>
- *Pérez-Salas, C., Ramos, C., Oliva, O., & Ortega, A. (2016). Bifactor modeling of the behavior rating inventory of executive function (BRIEF) in a chilean sample. *Perceptual and Motor Skills*, 1-20. <http://dx.doi.org/10.1177/0031512516650441>
- *Peters, C., Algina, J., Smith, S., & Daunic, A. (2012). Factorial validity of the behavior rating inventory of executive function (BRIEF) - teacher form. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 18(2), 168-18. <http://dx.doi.org/10.1080/09297049.2011.594427>
- *Piovesana, A., Ross, S., Whittingham, K., Ware, R., & Boyd, R. (2015). Stability of executive functioning measures in 8–17-year-old children with unilateral cerebral palsy. *The Clinical Neuropsychologist*, 1-19. <http://dx.doi.org/10.1080/13854046.2014.999125>
- *Pritchard, A., Kalback, S., McCurdy, M., & Capone, G. (2015). Executive functions among youth with down syndrome and co-existing neurobehavioural disorders. *Journal of Intellectual Disability Research*, 59(12), 1130-1141. <http://doi.org/10.1111/jir.12217>
- *Ransom, D., Bums, A., Youngstrom, E., Vaughan, C., Sady, M., & Gioia, G. (2016). Applying an evidence-based assessment model to identify students at risk for perceived academic problems following concussion. *Journal of the International Neuropsychological Society*, 22, 1038–1049. <https://doi.org/10.1017/S1355617716000916>
- *Rosenblum, S., Zandani, I., Deutsch-Castel, T., & Meyer, S. (2019). The child evaluation checklist (CHECK): a screening questionnaire for detecting daily functional “red flags” of underrecognized neurodevelopmental disorders among preschool children. *Occupational Therapy International*, 1-13. <https://doi.org/10.1155/2019/6891831>

- *Rosenblum, S., Frisch, C., Deutsch-Castel, T., & Josman, N. (2015). Daily functioning profile of children with attention deficit hyperactive disorder: a pilot study using an ecological assessment. *Neuropsychological Rehabilitation*, 25(3), 402-418. <http://dx.doi.org/10.1080/09602011.2014.940980>
- *Roy, A., Allain, P., Roulin, J., Fournet, N., & Gall, D. (2015). Ecological approach of executive functions using the behavioural assessment of the dysexecutive syndrome for children (BADSC-C): developmental and validity study. *Journal of Clinical and Experimental Neuropsychology*, 37(9), 956-971. <http://dx.doi.org/10.1080/13803395.2015.1072138>
- *Sadeh, S. Burns, M., & Sullivan, A. (2012). Examining an Executive Function Rating Scale as a Predictor of Achievement in Children at Risk for Behavior Problems. *School Psychology Quarterly*, 27(4), 236-246. <http://dx.doi.org/10.1037/spq0000012>
- Santana, A., Melo, M., & Minervino, C. (2019). Instrumentos de avaliação das funções executivas: revisão sistemática dos últimos cinco anos. *Avaliação Psicológica*, 18(1), 96-107. <http://dx.doi.org/10.15689/ap.2019.1801.14668.11>
- *Schneider, H., Ryan, M., & Mahone, E. (2019). Parent versus teacher ratings on the BRIEF-preschool version in children with and without ADHD. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 1-17. <https://doi.org/10.1080/09297049.2019.1617262>
- *Shum, K., Zheng, Q., Chak, G., Kei, K., Lam, C., Lam, I., Lok, C., & Tang, J. (2020). Dimensional structure of the BRIEF2 and its relations with ADHD symptoms and task performance on executive functions in chinese children. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 1-26. <https://doi.org/10.1080/09297049.2020.1817355>
- Souza, A., Alexandre, N., Guirardello, E. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Revista Epidemiologia e Serviços de Saúde*, 26(3), 649-659. <http://doi.org/10.5123/S1679-49742017000300022>

- *Spiegel, J., Lonigan, C., & Phillips, B. (2016). Factor structure and utility of the behavior rating inventory of executive function - preschool version. *Psychological Assessment*, 1-15. <http://dx.doi.org/10.1037/pas0000324>
- Spitzer, L., Binkofski, F., & Bruehl, S. (2019). Executive functions in aphasia: a novel aphasia screening for cognitive flexibility in everyday communication. *Neuropsychological Rehabilitation*, 1-20. <https://doi.org/10.1080/09602011.2019.1601572>
- *Stichter, J., Herzog, M., Owens, S., & Malugen, E. (2016). Manualization, feasibility, and effectiveness of the school – based social competence intervention for adolescents (SCI-A). *Psychology in the Schools*, 53(6), 583- 600. <http://doi.org/10.1002/pits.21928>
- *Strang, J., Anthony, L., Yerys, B., Hardy, K., Wallace, G., Armour, A., Dudley, K., Kenworthy, L. (2017). The flexibility scale: development and preliminary validation of a cognitive flexibility measure in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 1-17. <http://doi.org/10.1007/s10803-017-3152-y>
- *Sulik, M., & Obradovic, J. (2018). Teachers' rankings of children's executive functions: validating a methodology for school-based data collection. *Journal of Experimental Child Psychology*, 173, 136-154. <https://doi.org/10.1016/j.jecp.2018.01.016>
- *Tamm, L., & Peugh, J. (2018). Concordance of teacher-rated and performance-based measures of executive functioning in preschoolers. *Child Neuropsychology*, 1-16. <https://doi.org/10.1080/09297049.2018.1484085>
- Tamm, L., & Peugh, J. (2019). Concordance of teacher-rated and performance-based measures of executive functioning in preschoolers. *Child Neuropsychology*, 25(3), 410-424. <https://doi.org/10.1080/09297049.2018.1484085>
- *Tan, A., Delgaty, L., Steward, K., & Bunner, M. (2018). Performance-based measures and behavioral ratings of executive function in diagnosing attention-deficit/hyperactivity disorder in children. *ADHD Attention Deficit and Hyperactivity Disorders*, 1-8. <https://doi.org/10.1007/s12402-018-0256-y>

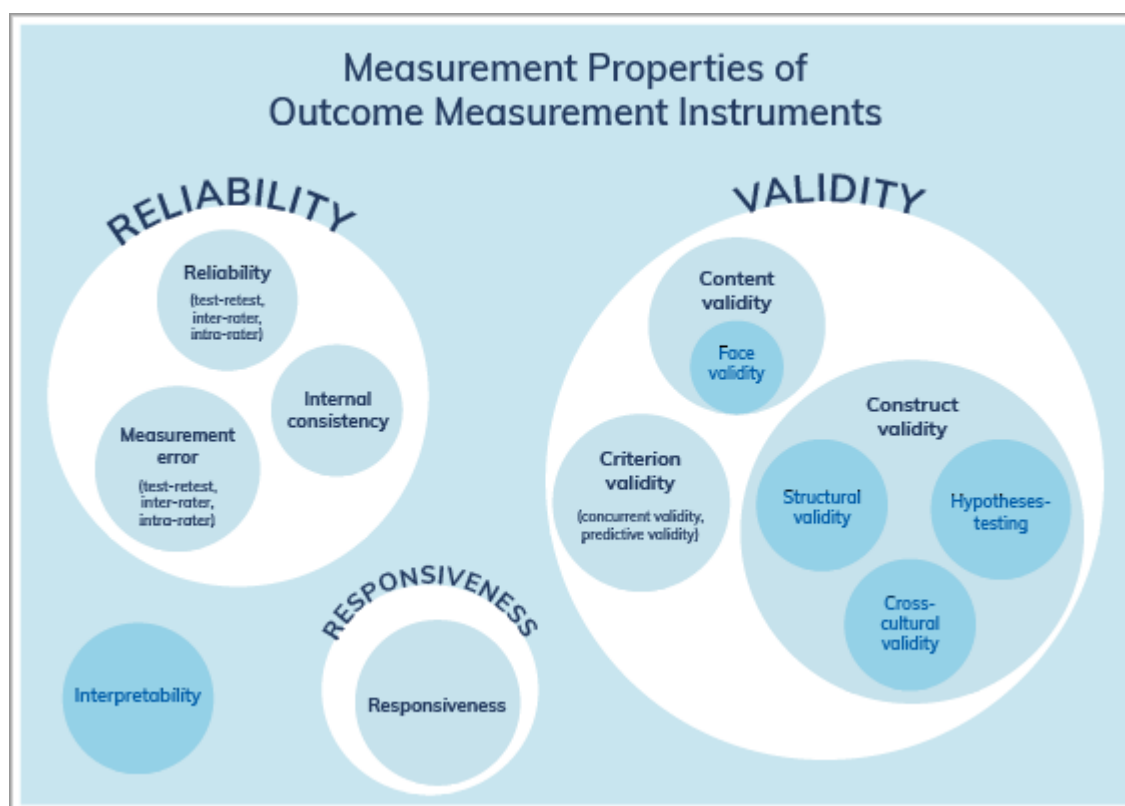
- Terwee, C.B., Mokkink, L.B., van Poppel, M.N.M. *et al.* (2010). Qualitative attributes and measurement properties of physical activity questionnaires. *Sports Med*, 40(7), 525–537. <https://doi.org/10.2165/11531370-000000000-00000>
- *Teunisse, J., Roelofs, R., Verhoeven, E., Cuppen, L., Mol, J., & Berger, H. (2012). Flexibility in children with autism spectrum disorders (ASD): inconsistency between neuropsychological tests and parent-based rating scales. *Journal of Clinical and Experimental Neuropsychology*, 34(7), 714-723. <http://dx.doi.org/10.1080/13803395.2012.670209>
- Thompson, A., & Steinbeis, N. (2020). Sensitive periods in executive function development. *Current Opinion in Behavioral Sciences*, 36, 98-105. <https://doi.org/10.1016/j.cobeha.2020.08.001>
- *Torske, T., Naerland, T., Oie, M., & Andreassen, O. (2018). Metacognitive aspects of executive function are highly associated with social functioning on parent-rated measures in children with autism spectrum disorder. *Frontiers in Behavioral Neuroscience*, 11(258), 1-12. <http://dx.doi.org/10.3389/fnbeh.2017.00258>
- *Usher, A., Leon, S., Stanford, L., Holmbeck, G., & Bryant, F. (2015). Confirmatory factor analysis of the behavior rating inventory of executive functioning (BRIEF) in children and adolescents with ADHD. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 1-14. <http://dx.doi.org/10.1080/09297049.2015.1060956>
- *Wang, L., Pollastri, A., Vujik, P., Hill, E., Lee, B., Samkavit, A., Braaten, E., Ablon, J., & Doyle, A. (2018). Reliability and validity of the thinking skills inventory, a screening tool for cross-diagnostic skill deficits underlying youth behavioral challenges. *Journal of Psychopathology and Behavioral Assessment*, 1-16. <https://doi.org/10.1007/s10862-018-9703-5>
- *Warnick, B., Drake, M., & Vidrine, S. (2015). Test Review. *Journal of Psychoeducational Assessment*, 33(6), 597-601. <http://doi.org/10.1177/0734282914564346>

Zelazo, P., & Muller, U. (2002). Executive function in typical and atypical development. *Blackwell Handbook of Childhood Cognitive Development*, 445-469.
<https://doi.org/10.1002/9780470996652.ch20>

As referências assinaladas com um asterisco (*) indicam quais os estudos foram incluídos neste estudo.

Anexos

Anexo 1. Taxonomia de propriedades de medida do Cosmin



Fonte: COnsensus-based Standards for the selection of health Measurement Instruments - 2006

Anexo 2. Método de Classificação das Propriedades de Medida

<i>Alfa de Cronbach</i>	<i>Kappa de Cohen</i>	Índice de Validade de Conteúdo (IVC)	Coefficiente de Correlação Interclasse (CCI)
> 0,90 – Muito Boa	>0,80 – Muito Boa	>0,90 – Muito Boa	>0,50 – Fraca
0,80 - 0,90 – Boa	0,61 – 0,80 – Boa	0,80 – Boa	0,50 – 0,75 – Moderada
0,70 – 0,80 – Razoável	0,41 – 0,60 – Moderada	<0,78 – Fraca	0,75 – 0,90 – Boa
0,60 – 0,70 – Fraca	0,21 – 0,40 – Razoável		<0,90 - Excelente
< 0,60 – Inadmissível	< 0,20 - Fraca		

Anexo 3. Lista de Verificação *Cosmin Risk of Bias*

Box 2. Content validity					
2a. Asking patients about relevance					
Design requirements		very good	adequate	doubtful	inadequate
1	Was an appropriate method used to ask patients whether each item is <u>relevant</u> for their experience with the condition?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Not clear if patients were asked whether <u>each</u> item is relevant or doubtful whether the method was appropriate	Method used not appropriate or patients not asked about the relevance of all items
2	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear	
3	Were skilled group moderators/interviewers used?	Skilled group moderators/interviewers used	Group moderators/interviewers had limited experience or were trained specifically for the study	Not clear if group moderators/interviewers were trained or group moderators/interviewers not trained and no experience	Not applicable
4	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide	Not applicable
5	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/interviews	No recording and no notes Not applicable
Analyses					
6	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate
7	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	

2b Asking patients about comprehensiveness					
Design requirements		very good	adequate	doubtful	inadequate
					NA
8	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate
9	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear	
10	Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience	Not applicable
11	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide	Not applicable
12	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes Not applicable
Analyses					
13	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate
14	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	

2c Asking patients about comprehensibility					
Design requirements		very good	adequate	doubtful	inadequate
15	Was an appropriate qualitative method used for assessing the <u>comprehensibility</u> of the PROM instructions, items, response options, and recall period?	Widely recognized or well justified qualitative method used	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was appropriate or not clear if patients were asked about the comprehensibility of the items, response options or recall period or patients not asked about the comprehensibility of the PROM instructions	Method used not appropriate or patients not asked about the comprehensibility of the items, response options, or recall period
16	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear	
17	Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience	
18	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide	Not applicable
19	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes Not applicable
Analyses					
20	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate
21	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	

2d. Asking professionals about relevance

Design requirements		very good	adequate	doubtful	inadequate	NA
22	Was an appropriate method used to ask professionals whether each item is <u>relevant</u> for the construct of interest?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Not clear if professionals were asked whether <u>each</u> item is relevant or doubtful whether the method was appropriate	Method used not appropriate or professionals not asked about the relevance of all items	
23	Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included or relevant professionals were not included		
24	Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
Analyses						
25	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	

26	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis		
----	---	---	--	--	--	--

2e. Asking professionals about comprehensiveness

Design requirement		very good	adequate	doubtful	inadequate	NA
27	Was an appropriate method used for assessing the <u>comprehensiveness</u> of the PROM?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
28	Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included or relevant professionals were not included		
29	Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	≥7 ≥50	4-6 ≥30	<4 or not clear <30 or not clear		
Analyses						
30	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	

31	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis or only one researcher involved in the analysis	
----	---	---	--	--	--

Box 3. Structural validity					
Does the scale consist of effect indicators, i.e. is it based on a reflective model? ¹ yes / no					
Does the study concern unidimensionality or structural validity? ² unidimensionality / structural validity					
Statistical methods	very good	adequate	doubtful	inadequate	NA
1 For CTT: Was exploratory or confirmatory factor analysis performed?	Confirmatory factor analysis performed	Exploratory factor analysis performed		No exploratory or confirmatory factor analysis performed	Not applicable
2 For IRT/Rasch: does the chosen model fit to the research question?	Chosen model fits well to the research question	Assumable that the chosen model fits well to the research question	Doubtful if the chosen model fits well to the research question	Chosen model does not fit to the research question	Not applicable
3 Was the sample size included in the analysis adequate?	FA: 7 times the number of items and ≥ 100	FA: at least 5 times the number of items and ≥ 100 ; OR at least 6 times number of items but < 100	FA: 5 times the number of items but < 100	FA: < 5 times the number of items	
	Rasch/1PL models: ≥ 200 subjects	Rasch/1PL models: 100-199 subjects	Rasch/1PL models: 50-99 subjects	Rasch/1PL models: < 50 subjects	
	2PL parametric IRT models OR Mokken scale analysis: ≥ 1000 subjects	2PL parametric IRT models OR Mokken scale analysis: 500-999 subjects	2PL parametric IRT models OR Mokken scale analysis: 250-499 subjects	2PL parametric IRT models OR Mokken scale analysis: < 250 subjects	

Other					
4	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws (e.g. rotation method not described)	Other important methodological flaws (e.g. inappropriate rotation method)

Box 4. Internal consistency					
Does the scale consist of effect indicators, i.e. is it based on a reflective model? ¹ yes / no					
<i>Design requirements</i>	very good	adequate	doubtful	inadequate	NA
1 Was an internal consistency statistic calculated for each unidimensional scale or subscale separately?	Internal consistency statistic calculated for each unidimensional scale or subscale		Unclear whether scale or sub scale is unidimensional	Internal consistency statistic NOT calculated for each unidimensional scale or sub scale	
<i>Statistical methods</i>					
2 For continuous scores: Was Cronbach's alpha or omega calculated?	Cronbach's alpha, or Omega calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated	Not applicable
3 For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated	Not applicable
4 For IRT-based scores: Was standard error of the theta (SE(θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	SE(θ) or reliability coefficient calculated			SE(θ) or reliability coefficient NOT calculated	Not applicable
<i>Other</i>					
5 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 6. Reliability					
<i>Design requirements</i>	very good	adequate	doubtful	inadequate	NA
1 Were patients stable in the interim period on the construct to be measured?	Evidence provided that patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate or time interval was not stated	Time interval NOT appropriate	
3 Were the test conditions similar for the measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar	
<i>Statistical methods</i>					
4 For continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated	Not applicable
5 For dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			No kappa calculated	Not applicable

6	For ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated or not described		Not applicable
7	For ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described			Not applicable
<i>Other</i>						
8	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 8. Criterion validity						
		very good	adequate	doubtful	inadequate	NA
Statistical methods						
1	For continuous scores: Were correlations, or the area under the receiver operating curve calculated?	Correlations or AUC calculated			Correlations or AUC NOT calculated	Not applicable
2	For dichotomous scores: Were sensitivity and specificity determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated	Not applicable
Other						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 10. Responsiveness						
10a. Criterion approach (i.e. comparison to a gold standard)						
		very good	adequate	doubtful	inadequate	NA
Statistical methods						
1	For continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	Correlations or Area under the ROC Curve (AUC) calculated			Correlations or AUC NOT calculated	Not applicable
2	For dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated			Sensitivity and specificity NOT calculated	Not applicable
Other						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)						
Design requirements		very good	adequate	doubtful	inadequate	NA
4	Is it clear what the comparator instrument(s) measure(s)?	Constructs measured by the comparator instrument(s) is clear			Constructs measured by the comparator instrument(s) is not clear	
5	Were the measurement properties of the comparator instrument(s) sufficient?	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s) OR evidence of poor quality of comparator instrument(s)	
Statistical methods						
6	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method were appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
7	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)						
Design requirements		very good	adequate	doubtful	inadequate	NA
8	Was an adequate description provided of important characteristics of the subgroups?	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor or no description of the important characteristics of the subgroups		
Statistical methods						
9	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
10	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10d. Construct approach: (i.e. hypotheses testing: before and after intervention)						
Design requirements		very good	adequate	doubtful	inadequate	NA
11	Was an adequate description provided of the intervention given?	Adequate description of the intervention		Poor description of the intervention	NO description of the intervention	
Statistical methods						
12	Was the statistical method appropriate for the hypotheses to be tested?	Statistical method was appropriate	Assumable that statistical method was appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other						
13	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	