



**SISTEMA DE  
ACREDITAÇÃO**

Provas e Testes  
Psicológicos

**SISTEMA DE  
ACREDITAÇÃO**  
Provas e Testes  
Psicológicos

**CRITÉRIOS E LINHAS  
ORIENTADORAS**  
— 2ª VERSÃO

# SISTEMA DE ACREDITAÇÃO DE PROVAS E TESTES PSICOLÓGICOS

## **DOCUMENTO BASE\*:**

Sistema de Acreditação de Provas e Testes Psicológicos: Critérios e Orientações para Submissão e Avaliação

## **DOCUMENTOS ASSOCIADOS:**

Sistema de Acreditação de Provas e Testes Psicológicos: Modelo de Submissão a Acreditação de Provas e Testes Psicológicos

## **SISTEMA DE ACREDITAÇÃO DE PROVAS E TESTES PSICOLÓGICOS:**

Modelo de Avaliação da Qualidade das Provas e Testes submetidos a Acreditação

## **AUTORES:**

Bártolo-Ribeiro, Rui  
Ferreira, Carla  
Gavela, João  
Machado, Magda  
Novo, Rosa  
Sá, José Silva  
Simões, Mário R.

## **DATA:**

Mai 2022

## **EDIÇÃO:**

Ordem dos Psicólogos Portugueses

## **VERSÃO:**

2ª versão- integra as recomendações aprovadas em Reunião de Direcção em 30/11/2023

## **RECOMENDAÇÃO DE CITAÇÃO DO DOCUMENTO BASE:**

Bártolo-Ribeiro, R., Ferreira, C., Gavela, J., Machado, M., Novo, R., Sá, J. S., & Simões, M. R. (2022). *Sistema de Acreditação de Provas e Testes Psicológicos: Critérios e Orientações para Submissão e Avaliação*. Ordem dos Psicólogos Portugueses.

---

\* Documento produzido pelo Grupo de Trabalho constituído a partir do Conselho Consultivo para a Acreditação de Testes e Provas Psicológicas da OPP (<https://www.ordemdospsicologos.pt/pt/comissoes/comissao/index/conselho-consultivo-para-a-acreditacao-de-testes-e-provas-psicologicas-da-opp>), e que, numa fase inicial de elaboração, contou com a participação de três outros elementos – Cristina Soeiro, Elizabete Reis e Rui Carreteiro.



Conforme Despacho Interno N.º 30/2018 da OPP, foi criado o Conselho Consultivo para a Acreditação de Testes e Provas Psicológicas, com a missão de dar apoio ao cumprimento das seguintes metas: “1 – Criação de um mecanismo de acreditação de Testes e Provas Psicológicas; 2 – Disponibilização de informação relativa aos Testes e outros instrumentos de avaliação psicológica e das condições da sua utilização de forma estruturada.”

Nesse sentido, os elementos constituintes deste Conselho cedo reconheceram a necessidade de partir de orientações internacionais de referência e modelos já existentes em outros países, e de adaptar e conceber orientações, parâmetros, indicadores e critérios de avaliação adequados à evolução, ocorrida e à realidade nacional. O grupo de trabalho formado a partir das reuniões iniciais do referido Conselho veio assim a desenvolver o presente documento – **Sistema de Acreditação de Provas e Testes Psicológicos: Critérios e Orientações para Submissão e Avaliação** – que visa dar resposta à solicitação da OPP e concretizar a adaptação, para a realidade portuguesa, do modelo de descrição e avaliação dos instrumentos de avaliação psicológica desenvolvido pela EFPA (Board of Assessment, 2013, *Review Model for the Description and Evaluation of Psychological and Educational Tests: Test Review Form. Version 4.2.6* (<http://www.efpa.eu/documents-/>)).

O presente documento teve igualmente em consideração a adaptação para outros países do modelo da EFPA, versões 4.2.6 e anteriores (e.g., Hernández et al., 2016), assim como o referencial de avaliação utilizado pelo Conselho Federal de Psicologia do Brasil (Diário Oficial da União, Resolução nº 9, de 25 de abril de 2018). Além destes documentos de referência, incorporou também as recomendações mais gerais constantes do *Standards for Educational and Psychological Testing da American Psychological Association* (APA, 2014) e da *International Test Commission* (e.g., ITC, 2017), bem como outros contributos (e.g., Van de Vijver & Poortinga, 2005).

**O Sistema de Acreditação de Provas e Testes Psicológicos: Critérios e Orientações para Submissão e Avaliação** constitui-se como um documento integrador que serve de referencial a outros dois documentos, destinados à operacionalização dos processos de submissão e de avaliação: o **Modelo de Submissão a Acreditação de Provas e Testes Psicológicos** e o **Modelo de Avaliação da Qualidade das Provas e Testes submetidos a Acreditação**. O primeiro documento tem por objectivo indicar e estruturar a informação técnica específica que deve ser facultada sobre a prova ou teste psicológico aquando do Pedido da sua Acreditação por parte da entidade requerente. O segundo documento visa sistematizar e orientar os processos avaliativos de cada prova ou teste psicológico de modo a sustentar o Parecer Final relativo à Acreditação.

A criação de um qualquer modelo de avaliação da qualidade de instrumentos de medida psicológica pode levantar importantes questões, nomeadamente ao nível da aplicabilidade dos indicadores e critérios considerados à diversidade das dimensões avaliadas e à multiplicidade de metodologias de desenvolvimento e de validação das medidas para diferentes objectivos e contextos avaliativos. Contudo, tais questões não devem colocar em causa o objectivo final de disponibilizar aos(as) psicólogos(as) informação estruturada sobre as qualidades exigíveis aos instrumentos de avaliação psicológica e às condições da sua utilização.

Por outro lado, é importante reconhecer a natureza dinâmica dos sistemas de avaliação da qualidade no âmbito da Avaliação Psicológica e a consequente necessidade de progressivamente incorporar, nos sistemas de acreditação, o sentido da evolução teórica, técnica e tecnológica no processo de avaliação. Deste modo, será feita uma revisão do presente modelo a cada 3 anos, ou antes, caso surja uma indicação teórica ou procedimental que altere os pressupostos aqui redigidos.

# ÍNDICE

<b>Nota Prévia</b>	<b>03</b>
<b>I. Identificação da Prova ou Teste Psicológico</b>	<b>08</b>
1. Nome e sigla da versão portuguesa do instrumento	08
2. Autor(es) da versão portuguesa do instrumento	08
3. Editor da versão portuguesa do instrumento	08
4. Data de publicação/edição da versão portuguesa do instrumento	08
5. Nome e sigla do instrumento na sua versão original	08
6. Autor(es) do instrumento na sua versão original	08
7. Editor do instrumento na sua versão original	08
8. Data de publicação/edição do instrumento na sua versão original	08
9. População-alvo	08
10. Administração	09
11. Condições de Administração	09
12. Tempo de aplicação	09
13. Formato dos itens	09
14. Número de itens	09
15. Área(s) de Aplicação	10
16. Nível de Qualificação Exigida	10
17. Constructo(s) avaliado(s)	10
18. Área(s) de Conteúdo	11
19. Requisitos do/a Respondente	11
<b>II. Avaliação da Prova ou Teste Psicológico</b>	<b>12</b>
<b>A. Materiais e documentação do instrumento (teste ou prova)</b>	<b>12</b>
1. Instrumentos de avaliação em suporte não digital	12
1.1 Materiais	12
1.2 Instruções de administração	12
1.3 Registo das respostas	12
1.4 Manual técnico	13
2. Instrumentos de avaliação em suporte digital	13
2.1 Software	13
2.2 Instruções de administração	13
2.3 Registo das respostas	13
2.4 Segurança contra acessos não autorizados	13
2.5 Manual técnico	14
Avaliação global dos materiais e documentação da prova ou teste psicológico	14
<b>B. Fundamentação teórica</b>	<b>14</b>
1. Fundamentos teóricos	14
2. Constructo(s)/Conceito(s)	14
3. Revisão da literatura científica relativa à(s) área(s) de aplicação	14
4. Revisão da literatura científica relativa à(s) área(s) de conteúdo	15
Avaliação global da fundamentação teórica	15
<b>C. Análise dos itens</b>	<b>15</b>
1. Redacção dos itens	15
2. Dados relativos à análise dos itens	15
2.1 Dificuldade e variabilidade dos itens	15
2.2 Discriminação dos itens	16
2.3 Análise dos distratores	16
Avaliação global da análise dos itens	16

<b>D. Fiabilidade/Precisão</b>	<b>17</b>
1. Indicadores de fiabilidade/precisão	17
2. Consistência interna	19
2.1 Qualidade da(s) amostra(s)	19
2.1.1 Natureza da(s) amostra(s)	19
2.1.2 Dimensão da(s) amostra(s)	19
2.2 Coeficiente(s) de consistência interna	20
2.2.1 Qualidade do(s) resultado(s)	20
3. Estabilidade temporal	20
3.1 Qualidade da(s) amostra(s)	20
3.1.1 Natureza da(s) amostra(s)	20
3.1.2 Dimensão da(s) amostra(s)	21
3.2 Qualidade do(s) resultado(s)	21
4. Equivalência das formas paralelas	21
4.1 Qualidade da(s) amostra(s)	21
4.1.1 Natureza da(s) amostra(s)	21
4.1.2 Dimensão da(s) amostra(s)	22
4.2 Qualidade do(s) resultado(s)	22
5. Fiabilidade/Precisão baseada na TRI	22
5.1 Qualidade da(s) amostra(s)	22
5.1.1 Natureza da(s) amostra(s)	22
5.1.2 Dimensão da(s) amostra(s)	23
5.2 Coeficiente(s) baseados na TRI	23
5.2.1 Qualidade do(s) resultado(s)	23
6. Acordo entre avaliadores	24
6.1 Qualidade da(s) amostra(s)	24
6.1.1 Natureza da(s) amostra(s)	24
6.1.2 Dimensão da(s) amostra(s)	24
6.2 Coeficiente(s) de acordo entre avaliadores	24
6.2.1 Qualidade do(s) resultado(s)	25
Avaliação global da fiabilidade/precisão	25
<b>E. Validade</b>	<b>25</b>
1. Evidências de validade	26
2. Evidências baseadas no conteúdo	27
2.1 Qualidade da representação do conteúdo ou domínio	27
3. Evidências baseadas na relação com critérios externos	27
3.1 Validade concorrente	27
3.1.1 Qualidade da(s) amostra(s)	27
a. Número e natureza da(s) amostra(s)	27
b. Dimensão da(s) amostra(s)	28
3.1.2 Qualidade dos critérios externos	28
3.1.3 Qualidade dos resultados	28
3.2 Validade preditiva	28
3.2.1 Qualidade da(s) amostra(s)	29
a. Número e natureza da(s) amostra(s)	29
b. Dimensão da(s) amostra(s)	29
3.2.2 Qualidade dos critérios externos	29
3.2.3 Qualidade dos resultados	30
3.3 Validade pós-ditiva	30
3.3.1 Qualidade da(s) amostra(s)	30
a. Número e natureza da(s) amostra(s)	30
b. Dimensão da(s) amostra(s)	31
3.3.2 Qualidade dos critérios externos	31
3.3.3 Qualidade dos resultados	31
4. Evidências baseadas na relação entre constructos	31
4.1 Validade convergente	31
4.1.1 Qualidade da(s) amostra(s)	31
a. Número e natureza da(s) amostra(s)	31
b. Dimensão da(s) amostra(s)	32
4.1.2 Qualidade das medidas utilizadas	32

4.1.3 Qualidade dos resultados	32
4.2 Validade discriminante	32
4.2.1 Qualidade da(s) amostra(s)	33
a. Número e natureza da(s) amostra(s)	33
b. Dimensão da(s) amostra(s)	33
4.2.2 Qualidade das medidas utilizadas	33
4.2.3 Qualidade dos resultados	33
4.3 Parâmetros relativos à estrutura interna	34
4.3.1 Qualidade da(s) amostra(s)	34
a. Número e natureza da(s) amostra(s)	34
b. Dimensão da(s) amostra(s)	34
4.3.2 Intercorrelações	34
a. Qualidade dos resultados	35
4.3.3 Análise fatorial	35
a. Metodologia e qualidade dos resultados	35
4.3.4 Aplicação da TRI	35
a. Metodologia e qualidade dos resultados	35
4.4 Parâmetros relativos à comparação de grupos	35
4.4.1 Diferenças entre grupos relevantes	36
a. Qualidade dos resultados	36
Avaliação global da validade	36
<b>F. Normas</b>	<b>36</b>
1. Tipos de Normas	37
2. Dados normativos	37
2.1 Critérios de selecção da(s) amostra(s)	37
2.2 Qualidade da(s) amostra(s)	37
2.2.1 Número e natureza da(s) amostra(s)	37
2.2.2 Dimensão da(s) amostra(s)	38
2.3 Actualidade dos dados normativos	38
2.4 Estatística descritiva dos resultados obtidos pelos grupos normativos	38
Avaliação global dos dados normativos	39
3. Critério(s) e/ou dados de referência	39
3.1 Qualidade da(s) amostra(s)	39
3.1.1 Número e natureza da(s) amostra(s)	39
3.1.2 Dimensão da(s) amostra(s)	39
3.2 Pontos de corte	39
3.2.1 Acordo entre peritos	40
a. Número de peritos e metodologia utilizada	40
b. Indicador(es) de acordo entre peritos	40
c. Grau de acordo entre peritos	40
3.2.2 Critérios de natureza empírica	40
a. Estudos realizados e indicadores utilizados	40
3.2.3 Actualidade dos estudos	41
Avaliação global do(s) critério(s) e/ou dados de referência	41
Avaliação global das normas	41
<b>G. Sistema de cotação e de interpretação</b>	<b>42</b>
1. Sistema de cotação	42
1.1 Procedimentos de cotação	42
1.2 Instruções de cotação	42
Avaliação global do sistema de cotação	43
2. Resultados finais	43
2.1 Tipo(s) de resultados	43
Avaliação global dos resultados finais	44
3. Sistema de interpretação	44
3.1 Interpretação referenciada à norma	44
3.2 Interpretação referenciada a um critério	44
3.3 Outro(s) tipo(s) de interpretação	45
3.4 Orientações para interpretação	45
Avaliação global do sistema de interpretação	45

4. Relatórios gerados automaticamente	45
4.1 Tipo de relatório	46
4.2 Suscetibilidade de modificação	46
4.3 Nível de complexidade	46
Avaliação global dos relatórios gerados automaticamente	46
Avaliação global do sistema de cotação e de interpretação	47
<b>III. Parecer Final</b>	<b>48</b>
<b>IV. Referências Bibliográficas</b>	<b>49</b>

# I. IDENTIFICAÇÃO DA PROVA OU TESTE PSICOLÓGICO

1. Nome e sigla da versão portuguesa do instrumento

2. Autor(es) da versão portuguesa do instrumento

3. Editor da versão portuguesa do instrumento

4. Data de publicação/edição da versão portuguesa do instrumento

5. Nome e sigla do instrumento na sua versão original | Caso a versão portuguesa seja uma adaptação

6. Autor(es) do instrumento na sua versão original | Caso a versão portuguesa seja uma adaptação

7. Editor do instrumento na sua versão original | Caso a versão portuguesa seja uma adaptação

8. Data de publicação/edição do instrumento na sua versão original | Caso a versão portuguesa seja uma adaptação

9. População-alvo | Indicar o(s) intervalo(s) de idades de administração e grupos específicos aos quais se destina o instrumento de avaliação



## 10. Administração

Suporte Não Digital\*

Suporte Digital

Suporte Não Digital\* e Digital

\*No caso da administração em suporte não digital, especificar qual o formato (papel-e-lápis, consola, aparelhos, etc.)

## 11. Condições de Administração

Individual

Colectiva

Individual e Colectiva

## 12. Tempo de aplicação | O tempo de preparação, administração e classificação/devolução deve ser descrito em minutos

Tempo de preparação

Tempo de administração

Tempo de classificação/  
devolução

## 13. Formato dos itens | Assinalar todas as opções que se aplicam. No caso da opção "Outro(s)", especificar qual ou quais são os outros formatos de resposta

Resposta construída

Resposta dicotómica (sim/  
não;verdadeiro/falso,etc.)

Resposta gradativa/tipo Likert

Adjectivos bipolares

Escolha múltipla

Outro(s)\*

\*Indicar qual ou quais:

## 14. Número de itens | Se o instrumento de avaliação incluir várias escalas ou subescalas, indicar o número de itens em cada uma das escalas ou subescalas

**15. Área(s) de Aplicação** | Assinalar todas as opções que se aplicam. Caso seja seleccionada mais do que uma área de aplicação, indicar a justificação do autor/editor no espaço reservado para o efeito. Se não houver uma definição clara, assinalar a opção "Outra(s)", referindo qual a área de aplicação mais apropriada de acordo com as informações fornecidas no manual

Psicologia Clínica e da Saúde	<input type="checkbox"/>	Sexologia	<input type="checkbox"/>	Coaching Psicológico	<input type="checkbox"/>
Neuropsicologia	<input type="checkbox"/>	Psicologia da Educação	<input type="checkbox"/>	Psicologia Comunitária	<input type="checkbox"/>
Psicogerontologia	<input type="checkbox"/>	Psicologia Vocacional e do Desenvol. de Carreira	<input type="checkbox"/>	Psicologia da Saúde Ocupacional	<input type="checkbox"/>
Psicologia do Desporto	<input type="checkbox"/>	Necessidades Educativas Especiais	<input type="checkbox"/>	Avaliação de Condutores e Cand. a condutores	<input type="checkbox"/>
Psicologia da Justiça	<input type="checkbox"/>	Intervenção Precoce	<input type="checkbox"/>	Avaliação do pessoal de vigilância	<input type="checkbox"/>
Psicoterapia	<input type="checkbox"/>	Psicologia do Trabalho, Social e das Organizações	<input type="checkbox"/>	Outra(s)*	<input type="checkbox"/>

\*Indicar qual ou quais:

Justificação do editor/autor(es) para a utilização em mais do que uma área de aplicação:

## 16. Nível de Qualificação Exigida

Profissional ou Investigador(a) sem formação específica em Psicologia (e.g., medicina, educação, recursos humanos)	<input type="checkbox"/>
Profissional ou Investigador(a) sem formação específica em Psicologia (e.g., medicina, educação, recursos humanos), mas com formação específica e credenciada sobre o instrumento	<input type="checkbox"/>
Psicólogo(a) ou Investigador(a) em Psicologia	<input type="checkbox"/>
Psicólogo(a) com formação específica na área de aplicação (e.g., formação académica na área de aplicação da prova)	<input type="checkbox"/>
Psicólogo(a) com formação específica na área de aplicação (e.g., especialização profissional na área de aplicação da prova)	<input type="checkbox"/>
Psicólogo(a) com formação específica e credenciada sobre o instrumento	<input type="checkbox"/>

## 17. Constructo(s) avaliado(s)

**18. Área(s) de Conteúdo** | Assinalar todas as opções que se aplicam. Se não houver uma definição clara, assinalar a opção "Outra(s)", referindo qual a área de conteúdo mais apropriada de acordo com as informações fornecidas no manual

Aptidões	<input type="checkbox"/>	Inteligência	<input type="checkbox"/>	Atitudes, Valores e Crenças	<input type="checkbox"/>
Motivação, Interesses	<input type="checkbox"/>	Desenvolvimento	<input type="checkbox"/>	Personalidade	<input type="checkbox"/>
Dimensões afectivas/ emocionais	<input type="checkbox"/>	Psicomotricidade	<input type="checkbox"/>	Dimensões e competências familiares	<input type="checkbox"/>
Saúde Mental/Psicopatologia	<input type="checkbox"/>	Estilos de resposta	<input type="checkbox"/>	Funções cognitivas	<input type="checkbox"/>
Outra(s)*	<input type="checkbox"/>				

\*Indicar qual ou quais:

**19. Requisitos do/a Respondente** | Assinalar todas as opções que se aplicam. Caso considere necessário que o requerente cumpra outros requisitos, assinalar a opção "Outra(s)"

Coordenação	<input type="checkbox"/>	Compreensão da Língua utilizada na prova	<input type="checkbox"/>	Coordenação visuo-motora	<input type="checkbox"/>
Outra(s)*	<input type="checkbox"/>				

\*Indicar qual ou quais:

VOLTAR AO ÍNDICE

# II. AVALIAÇÃO DA PROVA OU TESTE PSICOLÓGICO

## A. MATERIAIS E DOCUMENTAÇÃO DO INSTRUMENTO (TESTE OU PROVA)

Se o instrumento incluir administração em suporte digital e em suporte não digital é necessário avaliar os pontos 1 e 2

### 1. Instrumentos de avaliação em suporte não digital

#### 1.1 Materiais | cadernos de teste, folhas de resposta, cadernos de registo e outros materiais

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Impressão e apresentação de qualidade; materiais duradouros e resistentes; formato adequado, estímulos claros e sem elementos distractores.

#### 1.2 Instruções de administração

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Instruções uniformes e objectivas, claras, precisas e adequadas à população-alvo e/ou adaptadas a grupos específicos. Inclusão de itens de exemplo e de treino, se aplicável.

#### 1.3 Registo das respostas

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = O procedimento de registo é adequado à natureza da tarefa, intuitivo e o mais simples possível para minimizar erros. A informação sobre o modo de registo é clara, em linguagem comum e acessível.

## 1.4 Manual técnico

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Para além das instruções de aplicação e de cotação, bem como de procedimentos e orientações para análise e interpretação, o manual é de fácil utilização e inclui informação estruturada e explícita relativa a: fundamentos teóricos e objectivos do instrumento; população a que se destina; descrição dos materiais do teste ou prova; fiabilidade/precisão e validade; dados normativos ou de referência normativa; bibliografia relevante sobre o instrumento.

## 2. Instrumentos de avaliação em suporte digital

### 2.1 Software

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Software robusto (face a falhas do utilizador ou problemas de ligação à rede), de qualidade (com evidência de conformidade técnica e funcional), intuitivo e de fácil utilização. Estímulos e design adequados à natureza do instrumento.

### 2.2 Instruções de administração

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Instruções uniformes e objectivas, claras, precisas e adequadas à população-alvo e/ou adaptadas a grupos específicos. Inclusão de itens de exemplo e de treino, se aplicável.

### 2.3 Registo das respostas

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = O procedimento de registo é adequado à natureza da tarefa e o mais simples possível para minimizar erros. A informação sobre o modo de registo, por parte de quem aplica ou de quem responde é simples, clara e intuitiva.

### 2.4 Segurança contra acessos não autorizados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Software dispõe de sistema de protecção do teste (itens e/ ou respostas) mediante utilização de chave de acesso, password, codificação, etc.

## 2.5 Manual técnico

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Para além das instruções de administração e de cotação, bem como de procedimentos e orientações para análise e interpretação, o manual é de fácil utilização e inclui informação estruturada e explícita relativa a: fundamentos teóricos e objectivos do instrumento; população a que se destina; descrição dos materiais do teste ou prova; fiabilidade/precisão e validade; dados normativos ou de referência normativa; bibliografia relevante sobre o instrumento.

## Avaliação global dos materiais e documentação da prova ou teste psicológico

Integração e ponderação em função das pontuações atribuídas às diversas alíneas dos pontos 1 e/ou 2. Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar em risco uma adequada utilização do instrumento.

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

## B. FUNDAMENTAÇÃO TEÓRICA

### 1. Fundamentos teóricos

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Clareza e adequação dos fundamentos teóricos.

### 2. Constructo(s)/Conceito(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Especificidade, clareza e articulação do(s) constructo(s)/conceito(s).

### 3. Revisão da literatura científica relativa à(s) área(s) de aplicação

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Especificidade, actualidade e representatividade da literatura relativa à(s) área(s) de aplicação da prova/teste.

#### 4. Revisão da literatura científica relativa à(s) área(s) de conteúdo

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Especificidade, actualidade e representatividade da literatura relativa à(s) área(s) de conteúdo e objectivos da prova/teste.

#### Avaliação global da fundamentação teórica

Integração e ponderação em função das pontuações atribuídas aos diversos pontos desta secção. Deve ser considerada toda a informação disponibilizada, não só no manual da(o) prova/teste, como também em outras publicações referenciadas pelo autor/ editor.

Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar em risco uma adequada utilização do instrumento.

**1\***

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

\*No caso de avaliação "Insuficiente", justificar:

## C. ANÁLISE DOS ITENS

#### 1. Redacção dos itens

NÃO APLICÁVEL

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**3 (Bom)** = Os itens estão formulados de modo claro e alinhados com as opções de resposta.

#### 2. Dados relativos à análise dos itens

##### 2.1 Dificuldade e variabilidade dos itens

NÃO APLICÁVEL

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**3 (Bom)** = Apresentada informação detalhada e sustentada relativamente à dificuldade e variabilidade dos itens.

## 2.2 Discriminação dos itens

NÃO APLICÁVEL

**0**  
SEM INFORMAÇÃO

**1**  
INSUFICIENTE

**2**  
SUFICIENTE

**3**  
BOM

Regra orientadora:

**3 (Bom)** = Apresentada informação detalhada e sustentada relativamente à discriminação dos itens.

## 2.3 Análise dos distratores

NÃO APLICÁVEL

**0**  
SEM INFORMAÇÃO

**1**  
INSUFICIENTE

**2**  
SUFICIENTE

**3**  
BOM

Regra orientadora:

**3 (Bom)** = Apresentada informação detalhada e sustentada relativamente à análise aos itens distratores.

## Avaliação global da análise dos itens

Integração e ponderação em função das pontuações atribuídas nos diversos pontos desta secção.

Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar em risco uma adequada utilização do instrumento.

**1\***  
INSUFICIENTE

**2**  
SUFICIENTE

**3**  
BOM

\*No caso de avaliação "Insuficiente", justificar:



## D. FIABILIDADE/PRECISÃO

A fiabilidade/precisão está ligada à capacidade de as pontuações no teste garantirem, com um determinado grau de confiança, que os resultados estão libertos da variação do erro de medição. Corresponde à capacidade psicométrica de o instrumento de medida conseguir reproduzir os mesmos resultados em aplicações diferidas no tempo ou entre formas paralelas/equivalentes em condições iguais de aplicação e/ou a consistência do conjunto de itens avaliadores da mesma dimensão numa prova, podendo ser determinada de várias formas.

A análise da fiabilidade/precisão deverá tomar em consideração:

**1. Contexto da aplicação e o tipo de decisão envolvido:** a utilização de instrumentos para tomar decisões sobre grupos de pessoas (e.g., clima organizacional, factores psicossociais) e a sua utilização para fazer avaliações individuais (e.g., selecção de pessoas, diagnóstico clínico), devendo os requisitos de fiabilidade/precisão serem mais exigentes para este último.

**2. Nível de interpretação:** se as escalas são interpretadas só por si ou agregadas a outras escalas para uma escala composta. Neste último caso, a fiabilidade/precisão da escala compósita deve ser o foco principal de avaliação (e não as fiabilidades das escalas componentes).

**3. Natureza das dimensões em avaliação:** a avaliação de dimensões do grupo de máximo desempenho (e.g., inteligência, funções cognitivas) obriga a requisitos de fiabilidade/precisão mais exigentes que a avaliação de instrumentos de resposta típica (e.g., personalidade).

### 1. Indicadores de fiabilidade/precisão | Assinalar todos os indicadores de fiabilidade/precisão apresentados na documentação do instrumento

Consistência interna (avaliar o ponto 2)

Estabilidade temporal (avaliar o ponto 3)

Equivalência das formas paralelas (avaliar o ponto 4)

Fiabilidade/Precisão baseada na TRI (avaliar o ponto 5)

Acordo entre avaliadores (avaliar o ponto 6)

Outro(s) procedimento(s)

\*Indicar qual ou quais:

Não são apresentados indicadores de fiabilidade/precisão (avançar para a “Avaliação global da fiabilidade/precisão”)

Especificação da(s) amostra(s) utilizada(s) para análise desse(s) indicador(es) de fiabilidade/precisão

**Um único coeficiente de fiabilidade/precisão**

Especificar a(s) escala(s) ou subescala(s), se aplicável:

**Diferentes tipos de coeficientes de fiabilidade/precisão**

Especificar a(s) escala(s) ou subescala(s), se aplicável:

**Coefficiente(s) de fiabilidade/precisão para grupos específicos de pessoas**

Especificar a(s) escala(s) ou subescala(s), se aplicável:

**Um único erro-padrão de medida**

Especificar a(s) escala(s) ou subescala(s), se aplicável:

**Erro-padrão de medida para grupos específicos de pessoas**

Especificar a(s) escala(s) ou subescala(s), se aplicável:

**Quantificação do erro de medida através da TRI**

Especificar a(s) escala(s) ou subescala(s), se aplicável:

**Outras informações**

Indicar qual:

## 2. Consistência interna

O uso de coeficientes de consistência interna é adequado para avaliar a fiabilidade/precisão da generalidade dos instrumentos, com excepção dos testes de velocidade, de escalas heterogéneas (também mencionadas como escalas empíricas ou de critérios-chave (Cronbach, 1970), de indicadores de efeito (Nunnally & Bernstein, 1994) e de traços emergentes (Schneider & Hough, 1995). O uso da consistência interna em escalas ipsativas pode enviesar a estimativa dos coeficientes de fiabilidade/precisão. As formas paralelas e o teste-reteste são métodos mais adequados para o estudo da fiabilidade/precisão desse tipo de escalas.

Os coeficientes de consistência interna calculados pela fórmula do alfa de Cronbach ou da análise factorial, por exemplo, proporcionam uma melhor estimativa da fiabilidade/precisão do que os coeficientes obtidos pela divisão em metades corrigidos com a fórmula de Spearman-Brown. Assim, o cálculo da fiabilidade/precisão com o uso da divisão em metades só é justificado se, por qualquer motivo, a informação sobre as respostas a cada um dos itens não estiver disponível.

### Adequação do método ao instrumento

Adequado

Não adequado (avançar para o ponto 3)

### 2.1 Qualidade da(s) amostra(s)

#### 2.1.1 Natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INADEQUADA(S)

**2**

ADEQUADA(S)

#### 2.1.2 Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

#### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $N < 200$

**Pontuação 2 (Suficiente)** =  $200 \leq N < 400$

**Pontuação 3 (Bom)** =  $N \geq 400$

#### Observações

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e a população-alvo à qual o instrumento se destina. A dimensão de amostras de grupos específicos terá em consideração a frequência ou a prevalência na população-alvo (e.g., grupos clínicos, faixas etárias, grupos minoritários).
- A avaliação deste parâmetro deve atender também ao número de estudos e de amostras utilizadas (natureza e dimensão).

## 2.2 Coeficiente(s) de consistência interna | Assinalar a(s) opção/ões aplicável/eis

Alfa de Cronbach ou KR-20

Guttman's Lambda-2

Omega (Análise Fatorial)

Theta (Análise Fatorial)

Outro(s)

Indicar qual ou quais:

Sem informação

### 2.2.1 Qualidade do(s) resultado(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

Regra orientadora (tomando em consideração a categoria do instrumento):

**Pontuação 1 (Insuficiente)** = Provas de Máximo Desempenho: ( $r < 0,70$ ); Provas de Resposta Típica: ( $r < 0,60$ )

**Pontuação 2 (Suficiente)** = Provas de Máximo Desempenho: ( $0,70 \leq r < 0,85$ ); Provas de Resposta Típica: ( $0,60 \leq r < 0,80$ )

**Pontuação 3 (Bom)** = Provas de Máximo Desempenho: ( $r \geq 0,85$ ); Provas de Resposta Típica: ( $r \geq 0,80$ )

## 3. Estabilidade temporal

A fiabilidade/precisão baseada em estudos de estabilidade temporal corresponde aos resultados da administração do mesmo instrumento à mesma amostra, nas mesmas condições, com um determinado intervalo de tempo. O cálculo da fiabilidade/precisão através de estudos de estabilidade temporal não é adequado para instrumentos de medida de variáveis tipo estado. Por outro lado, o intervalo de tempo entre as aplicações pode influenciar a magnitude do coeficiente de correlação.

Adequação do método ao instrumento

Adequado

Não adequado (avançar para o ponto 4)

### 3.1 Qualidade da(s) amostra(s)

#### 3.1.1 Natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INADEQUADA(S)

**2**

ADEQUADA(S)

### 3.1.2 Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $N < 100$

**Pontuação 2 (Suficiente)** =  $100 \leq N < 200$

**Pontuação 3 (Bom)** =  $N \geq 200$

**Observações**

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e a população-alvo à qual o instrumento se destina. A dimensão de amostras de grupos específicos terá em consideração a frequência ou a prevalência na população-alvo (e.g., grupos clínicos, faixas etárias, grupos minoritários).
- A avaliação deste parâmetro deve atender também ao número de estudos e de amostras utilizadas (natureza e dimensão).

### 3.2 Qualidade do(s) resultado(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora (tomando em consideração a categoria do instrumento):**

**Pontuação 1 (Insuficiente)** = Provas de Máximo Desempenho: ( $r < 0,70$ ); Provas de Resposta Típica: ( $r < 0,60$ )

**Pontuação 2 (Suficiente)** = Provas de Máximo Desempenho: ( $0,70 \leq r < 0,85$ ); Provas de Resposta Típica: ( $0,60 \leq r < 0,80$ )

**Pontuação 3 (Bom)** = Provas de Máximo Desempenho: ( $r \geq 0,85$ ); Provas de Resposta Típica: ( $r \geq 0,80$ )

## 4. Equivalência das formas paralelas

A estimativa da fiabilidade/precisão com estudos baseados na equivalência das formas paralelas/alternadas de um teste corresponde à administração das várias formas do instrumento a uma mesma amostra, nas mesmas condições, com um determinado intervalo de tempo. O intervalo de tempo entre aplicações deve tomar em consideração a familiaridade dos avaliados com as provas, o que pode influenciar a magnitude do coeficiente de correlação obtido.

**Adequação do método ao instrumento**

Adequado

Não adequado (avançar para o ponto 5)

### 4.1 Qualidade da(s) amostra(s)

#### 4.1.1 Natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INADEQUADA(S)

**2**

ADEQUADA(S)

#### 4.1.2 Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $N < 100$

**Pontuação 2 (Suficiente)** =  $100 \leq N < 200$

**Pontuação 3 (Bom)** =  $N \geq 200$

**Observações**

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e a população-alvo à qual o instrumento se destina. A dimensão de amostras de grupos específicos terá em consideração a frequência ou a prevalência na população-alvo (e.g., grupos clínicos, faixas etárias, grupos minoritários).
- A avaliação deste parâmetro deve atender também ao número de estudos e de amostras utilizadas (natureza e dimensão).

#### 4.2 Qualidade do(s) resultado(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

As duas formas do teste podem ser consideradas equivalentes se, para o mesmo grupo:

- os valores médios e as variâncias forem idênticas;
- as correlações entre as formas forem elevadas:

**Pontuação 1 (Insuficiente)** =  $r < 0,70$

**Pontuação 2 (Suficiente)** =  $0,70 \leq r < 0,85$

**Pontuação 3 (Bom)** =  $r \geq 0,85$

### 5. Fiabilidade/Precisão baseada na TRI

A avaliação da fiabilidade/precisão baseada na TRI, como, por exemplo, através da função de informação, está associada à precisão da estimativa do nível do traço individual. Na estimativa da fiabilidade/precisão pela TRI deve ser tomado em consideração o tipo de formato de resposta dos itens (e.g., dicotómicos, politómicos) e o modelo usado (1, 2 ou 3 parâmetros).

#### Adequação do método ao instrumento

Adequado

Não adequado (avançar para o ponto 6)

#### 5.1 Qualidade da(s) amostra(s)

##### 5.1.1 Natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INADEQUADA(S)

**2**

ADEQUADA(S)

### 5.1.2 Dimensão da(s) amostra(s) | A dimensão da(s) amostra(s) depende do tipo/escala de resposta aos itens e do modelo utilizado

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora tomando em consideração o número de parâmetros (Parshall, Davey, Spray, & Kalohn, 2002):**

**Pontuação 1 (Insuficiente)** = 1PL: < 200; 2PL: < 400; 3PL: < 700

**Pontuação 2 (Suficiente)** = 200 ≤ 1PL < 300; 400 ≤ 2PL ≤ 600; 700 ≤ 3PL < 1050

**Pontuação 3 (Bom)** = 1PL: ≥ 300; 2PL: ≥ 600; 3PL: ≥ 1050

**Observações:**

- O tamanho da amostra apresentado na regra orientadora aplica-se a modelos dicotômicos. Quando aplicados a modelos politômicos, os tamanhos de amostra podem ser menores (EFPA, 2013; Evers et al., 2013).
- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e a população-alvo à qual o instrumento se destina. A dimensão de amostras de grupos específicos terá em consideração a frequência ou a prevalência na população-alvo (e.g., grupos clínicos, faixas etárias, grupos minoritários).
- A avaliação deste parâmetro deve atender também ao número de estudos e de amostras utilizadas (natureza e dimensão).

### 5.2 Coeficiente(s) baseados na TRI | Assinalar a(s) opção/ões aplicável/eis

Rho

Fiabilidade/precisão das estimativas do Traço Latente

Função de informação

Outro(s) procedimento(s)

Indicar qual ou quais:

Sem informação

### 5.2.1 Qualidade do(s) resultado(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $r < 0,70$ ; Função de informação < 3,33

**Pontuação 2 (Suficiente)** =  $0,70 \leq r < 0,85$ ;  $3,33 \leq$  Função de informação < 7,50

**Pontuação 3 (Bom)** =  $r \geq 0,85$ ; Função de informação  $\geq 7,50$

**Observações:**

A regra orientadora para avaliação dos valores da função de informação do teste assume que esta é o somatório das funções de informação dos itens (Lord, 1980), e baseia-se na assunção de que a informação é  $1 / EP^2$ , e assumindo que  $r = 1 - EP^2$ . Chama-se a atenção de que os valores do EP (Erro-Padrão) e da Informação dependem do valor do traço latente e que cada teste tem um intervalo dentro qual o valor da informação é óptimo. A classificação não deve, a priori, basear-se nesse valor óptimo, mas no valor da informação relativo ao intervalo de pontuações que são de importância crítica. Se não forem referidas essas pontuações, a classificação deve ser baseada em o valor médio da informação (ver Reise & Havilund, 2005). Porque não há muita experiência com a utilização das linhas orientadoras apresentadas, aconselha-se precaução no seu uso (EFPA, 2013).

## 6. Acordo entre avaliadores

O acordo entre avaliadores deve ser considerado para instrumentos que requeiram maior julgamento da parte do avaliador na aplicação das regras de codificação e de cotação.

### Adequação do método ao instrumento

Adequado

Não adequado (avançar para o ponto 6)

## 6.1 Qualidade da(s) amostra(s)

### 6.1.1 Natureza da(s) amostra(s)

**0**  
SEM INFORMAÇÃO

**1**  
INADEQUADA(S)

**2**  
ADEQUADA(S)

### 6.1.2 Dimensão da(s) amostra(s) | A dimensão da(s) amostra(s) depende do tipo/escala de resposta aos itens e do modelo utilizado

**0**  
SEM INFORMAÇÃO

**1**  
INSUFICIENTE

**2**  
SUFICIENTE

**3**  
BOM

#### Regra orientadora:

Pontuação 1 (Insuficiente) =  $N < 50$

Pontuação 2 (Suficiente) =  $50 \leq N < 100$

Pontuação 3 (Bom) =  $N \geq 100$

#### Observações:

• A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e o número de itens que o constituem e a população-alvo à qual o instrumento se destina, o constructo examinado, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

## 6.2 Coeficiente(s) de acordo entre avaliadores | Assinalar a(s) opção/ões aplicável/eis

Percentagem de acordo

Coeficiente Kappa

Coeficiente de correlação intraclassa (ICC)

Coeficiente baseado na Teoria da Generalização

Outro(s)

Indicar qual ou quais:

Sem informação



## 6.2.1 Qualidade do(s) resultado(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

Regra orientadora:

Pontuação 1 (Insuficiente) = coeficiente < 0,55

Pontuação 2 (Suficiente) =  $0,55 \leq$  coeficiente < 0,75

Pontuação 3 (Bom) = coeficiente  $\geq$  0,75

## Avaliação global da fiabilidade/precisão

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção.

Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar em risco uma adequada utilização do instrumento.

**1\***

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

\*No caso de avaliação "Insuficiente", justificar:

## E. VALIDADE

Enquanto conceito, a validade traduz a grau ou extensão em que os resultados obtidos num instrumento medem, efectivamente, o constructo que esse instrumento pretende avaliar. A validade de um instrumento remete, assim, para as interpretações dos dados disponíveis e dos resultados obtidos nesse instrumento. E a fiabilidade/precisão, abordada no capítulo anterior, constitui um requisito da validade.

No processo de validação, o que é validado são as inferências realizadas a partir dos resultados obtidos com o teste, e não o teste propriamente dito. Tradicionalmente são considerados diferentes tipos de validade, os quais traduzem diferentes fontes de evidências de validade. A importância de uma ou outra evidência dependerá da natureza do instrumento e da utilização que se pretenda fazer dele.

As evidências mais relevantes baseiam-se:

(i) **no conteúdo;**

(ii) **na relação com critérios externos** (e.g., de natureza clínica, educacional) ou dados referentes a outros testes que avaliem o mesmo constructo; em qualquer dos casos pode constituir-se como:

- indicador de validade concorrente (quando medidas do teste e do critério são obtidas no mesmo período temporal);
- indicador de validade preditiva (quando medidas do teste antecedem temporalmente as medidas do critério);
- indicador de validade pós-ditiva (quando medidas do teste são obtidas posteriormente às medidas do critério).

(iii) **na relação entre constructos**, que pode constituir-se como:

- indicador de validade convergente (baseada na relação do constructo em análise com medidas de constructos relacionados);
- indicador de validade discriminante (baseada na ausência de relação do constructo em análise com medidas de constructos independentes);
- indicador da estrutura/organização interna (e.g., avaliação da estrutura factorial do instrumento).

**1. Evidências de validade** | Assinalar todas as evidências de validade apresentadas na documentação relativa ao instrumento

**Evidências baseadas no conteúdo (avaliar o ponto 2)**

**Evidências baseadas na relação com critérios externos**

**Validade concorrente (avaliar o ponto 3.1)**

**Validade preditiva (avaliar o ponto 3.2)**

**Validade pós-ditiva (avaliar o ponto 3.3)**

**Evidências baseadas na relação entre constructos**

**Validade convergente (avaliar o ponto 4.1)**

**Validade discriminante (avaliar o ponto 4.2)**

**Parâmetros relativos à estrutura interna** | avaliar o ponto 4.3.1, seguido do(s) pontos 4.3.2, 4.3.3 e/ou 4.3.4)

**Intercorrelações (avaliar o ponto 4.3.2)**

**Análise fatorial (avaliar o ponto 4.3.3)**

**Aplicação da TRI (avaliar o ponto 4.3.4)**

**Parâmetros relativos à comparação de grupos**

**Diferenças entre grupos relevantes (avaliar o ponto 4.4.1)**

**Outro(s) tipo(s) de evidências**

Indicar qual ou quais:

**Não são apresentadas evidências de validade (avançar para a “Avaliação global da validade”)**

## 2. Evidências baseadas no conteúdo

A validade de conteúdo caracteriza a extensão ou grau com que os itens de um instrumento avaliam uma amostra representativa da totalidade do conteúdo do domínio, constructo, desempenho ou comportamento-alvo. Por exemplo, num teste de avaliação de conhecimentos de aritmética, elaborado para o 3.º ano de escolaridade, é importante que o teste inclua itens que avaliem todas as operações de aritmética possíveis (adição, subtração, multiplicação, divisão).

### 2.1 Qualidade da representação do conteúdo ou domínio

0

SEM INFORMAÇÃO

1

INSUFICIENTE

2

SUFICIENTE

3

BOM

#### Regra orientadora:

**Pontuação 1 (Insuficiente)** = A evidência resultante de revisão sistemática da literatura científica relativa ao constructo e/ou referência significativa relativa a outros testes/instrumentos relevantes que avaliam o referido constructo é apresentada de forma lacunar e imprecisa.

**Pontuação 2 (Suficiente)** = É apresentada evidência resultante de revisão sistemática da literatura científica relativa ao constructo e/ou referência significativa relativa a outros testes/instrumentos relevantes que avaliam o referido constructo.

**Pontuação 3 (Bom)** = É apresentada uma definição precisa do domínio. Os itens/estímulos representam adequadamente todas as facetas do domínio avaliado. É apresentada evidência da validade de conteúdo relativa à versão final do teste [e.g., no caso de se tratar de um novo instrumento, e não de uma adaptação, são de valorizar a evidência documentada de consulta formal/estruturada, através de um procedimento sistematizado, com recurso a peritos (se aplicável)], bem como a evidência resultante de revisão sistemática da literatura científica relativa ao constructo e/ou referência significativa relativa a outros testes/instrumentos que avaliam o referido constructo. Quando se tratar de um novo instrumento, o recurso a peritos deve ser documentado considerando o grau de acordo relativamente à relevância de cada item, por exemplo, utilizando o ratio de validade de conteúdo de Lawshe (1975), cujos valores mínimos são:

- 7 ou menos peritos: 0,99

- 8 peritos: 0,75

- 9 peritos: 0,78

- 10 peritos: 0,62

#### Observações:

• Aspecto especialmente importante nos testes referidos ao critério e, particularmente, nos testes de rendimento escolar/académico. Avaliação sobre a extensão/qualidade da representação do conteúdo ou do domínio, através dos itens/estímulos. A documentação apresentada inclui avaliações de peritos (se aplicável) e revisão da literatura.

## 3. Evidências baseadas na relação com critérios externos

### 3.1 Qualidade da(s) amostra(s)

A validade concorrente é um tipo de validade relativa a critérios externos. Analisa o grau de relação entre os resultados no/a teste/prova e critérios externos (i.e., medida ou critério conhecidos do mesmo constructo ou de constructos similares, que são considerados aceitáveis). O intervalo temporal entre a administração do/a teste/prova e a medida deste critério externo é reduzido.

#### 3.1.1 Qualidade da(s) amostra(s)

##### a. Número e natureza da(s) amostra(s)

0

SEM INFORMAÇÃO

1

INSUFICIENTE

2

SUFICIENTE

3

BOM

#### Regra orientadora:

**Pontuação 1 (Insuficiente)** = Amostra(s) não relevante(s) para a medida do constructo na população-alvo

**Pontuação 2 (Suficiente)** = Uma amostra relevante para a medida do constructo na população-alvo

**Pontuação 3 (Bom)** = Duas ou mais amostras relevantes para a medida do constructo na população-alvo

#### Observações:

• O número e a natureza da(s) amostra(s) devem atender aos grupos específicos aos quais o instrumento se destina.

### b. Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

#### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $N < 75$  Pontuação

**2 (Suficiente)** =  $75 \leq N < 150$  Pontuação

**3 (Bom)** =  $N \geq 150$

#### Observações:

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e o número de itens que o constituem, a população-alvo à qual o instrumento se destina, o constructo examinado, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congêneres.

### 3.1.2 Qualidade dos critérios externos

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

#### Regra orientadora:

**Pontuação 1 (Insuficiente)** = O(s) critério(s) externo(s) utilizado(s) para a avaliação da validade concorrente não é/são representativo(s) do constructo.

**Pontuação 2 (Suficiente)** = Adequada selecção do(s) critério(s) externo(s) utilizado(s) para a avaliação da validade concorrente e as suas propriedades psicométricas são, pelo menos, satisfatórias.

**Pontuação 3 (Bom)** = Critério(s) externo(s) (e.g., clínicos, comportamentais, de desempenho e/ou instrumentos representativos das dimensões e com boas propriedades psicométricas, etc.) objectivo(s) e bem estabelecido(s) (na literatura teórica e/ou empiricamente).

### 3.1.3 Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

#### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $r < 0,55$

**Pontuação 2 (Suficiente)** =  $0,55 \leq r < 0,65$

**Pontuação 3 (Bom)** =  $r \geq 0,65$

### 3.2 Validade preditiva

A validade preditiva constitui uma outra evidência de validade relativa a critérios externos. Indica o grau em que os resultados no o/a teste/prova pode prever um desempenho ou comportamento futuros. É obtida a partir do exame da correlação entre o/a teste/prova e um critério externo reconhecido.

### 3.2.1 Qualidade da(s) amostra(s)

#### a. Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = Amostra(s) não relevante(s) da população-alvo.

**Pontuação 2 (Suficiente)** = Uma amostra relevante da população-alvo.

**Pontuação 3 (Bom)** = Duas ou mais amostras relevantes para a medida do constructo na população-alvo.

**Observações:**

- O número e a natureza da(s) amostra(s) devem atender aos grupos específicos aos quais o instrumento se destina, à literatura e à investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

#### b. Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $N < 50$

**Pontuação 2 (Suficiente)** =  $50 \leq N < 100$

**Pontuação 3 (Bom)** =  $N \geq 100$

**Observações:**

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e o número de itens que o constituem, a população-alvo à qual o instrumento se destina, o constructo examinado, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

### 3.2.2 Qualidade dos critérios externos

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = O(s) critério(s) externo(s) utilizado(s) para a avaliação da validade preditiva não é/são representativo(s) do constructo.

**Pontuação 2 (Suficiente)** = Adequada selecção do(s) critério(s) externo(s) utilizado(s) para a avaliação da validade preditiva e as suas propriedades psicométricas são, pelo menos, satisfatórias.

**Pontuação 3 (Bom)** = Critério(s) externo(s) objectivo(s) e bem estabelecido(s) (na literatura e/ou empiricamente); intervalos temporais entre a administração do instrumento e o(s) critério(s) externo(s) (i.e., preditor(es)) considerados adequados à natureza do instrumento.

### 3.2.3 Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $r < 0,20$

**Pontuação 2 (Suficiente)** =  $0,20 \leq r < 0,45$

**Pontuação 3 (Bom)** =  $r > 0,45$

**Observações:**

- Os valores a considerar deverão ser diretos, sem recurso à transformação da distribuição de resultados para correcção da restrição da amostra.
- Estes valores devem ser analisados considerando o contexto de utilização do instrumento, o intervalo de tempo entre a avaliação do(s) preditor(es) e o(s) critério(s) e a dimensão da(s) amostra(s). Por exemplo, enquanto um coeficiente 0.3 pode ser considerado bom para um contexto de selecção, em contexto escolar esse valor, ou mesmo um superior, pode ser mais facilmente obtido, e por isso não ter o mesmo significado.

### 3.3 Validade pós-ditiva

Os estudos de validade pós-ditiva são raros, abrangendo, sobretudo, estudos nas áreas educativa, forense e organizacional. A validade pós-ditiva constitui uma outra evidência de validade relativa a critérios externos. Relaciona resultados no/a teste/prova com evidências prévias existentes, relativas a ocorrências demonstrativas do mesmo constructo ou tipo de comportamento. Por exemplo: relação entre resultado em instrumento de medida da tendência de risco de acidentes e dados ou registos do mesmo tipo existentes em arquivos médicos.

#### 3.3.1 Qualidade da(s) amostra(s)

##### a. Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = Amostra(s) não relevante(s) para a medida do constructo na população-alvo

**Pontuação 2 (Suficiente)** = Uma amostra relevante para a medida do constructo na população-alvo

**Pontuação 3 (Bom)** = Duas ou mais amostras relevantes para a medida do constructo na população-alvo

**Observações:**

- O número e a natureza da(s) amostra(s) devem atender aos grupos específicos aos quais o instrumento se destina, à literatura e à investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

##### b. Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $N < 50$

**Pontuação 2 (Suficiente)** =  $50 \leq N < 100$

**Pontuação 3 (Bom)** =  $N \geq 100$

**Observações:**

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e o número de itens que o constituem, a população-alvo à qual o instrumento se destina, o constructo examinado, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

### 3.3.2 Qualidade dos critérios externos

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = O(s) critério(s) externo(s) utilizado(s) para avaliação da validade pós-ditiva não é/são representativo(s) do constructo.

**Pontuação 2 (Suficiente)** = Adequada selecção dos critérios externos para a avaliação da validade pós-ditiva e as suas propriedades psicométricas são, pelo menos, satisfatórias.

**Pontuação 3 (Bom)** = Critério(s) externo(s) objectivo(s) e bem estabelecido(s) (na literatura e/ou empiricamente); intervalos temporais entre a administração do instrumento e o(s) critério(s) externo(s) (i.e., preditor(es)) considerados adequados à natureza do instrumento.

### 3.3.3 Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $r < 0,20$

**Pontuação 2 (Suficiente)** =  $0,20 \leq r < 0,45$

**Pontuação 3 (Bom)** =  $r > 0,45$

**Observações:**

- Os valores a considerar deverão ser directos, sem recurso à transformação da distribuição de resultados para correcção da restrição da amostra.
- Estes valores devem ser analisados considerando o contexto de utilização do instrumento, o intervalo de tempo entre a avaliação do(s) preditor(es) e o(s) critério(s) e a dimensão da(s) amostra(s).

## 4. Evidências baseadas na relação entre constructos

### 4.1 Validade convergente

O teste ou instrumento que se pretende acreditar deve apresentar correlações elevadas com outros testes ou medidas que avaliam conceitos relacionados.

#### 4.1.1 Qualidade da(s) amostra(s)

##### a. Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

## b. Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $N < 75$

**Pontuação 2 (Suficiente)** =  $75 \leq N < 150$

**Pontuação 3 (Bom)** =  $N \geq 150$

### Observações:

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e o número de itens que o constituem, a população-alvo à qual o instrumento se destina, o constructo examinado, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

## 4.1.2 Qualidade das medidas utilizadas

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

### Regra orientadora:

**Pontuação 1 (Insuficiente)** = O(s) teste(s) marcador(es) utilizado(s) para a avaliação da validade convergente e as suas propriedades psicométricas não é/não são representativo(s) do constructo.

**Pontuação 2 (Suficiente)** = Adequada selecção do(s) teste(s) marcador(es) utilizado(s) para a avaliação da validade convergente e as suas propriedades psicométricas são, pelo menos, satisfatórias.

**Pontuação 3 (Bom)** = Critérios (e.g., clínicos, comportamentais, de desempenho e/ou instrumentos representativos das dimensões e com boas propriedades psicométricas, etc.) objectivos e bem estabelecidos (na literatura teórica e/ou empiricamente).

## 4.1.3 Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $r < 0,55$

**Pontuação 2 (Suficiente)** =  $0,55 \leq r < 0,65$

**Pontuação 3 (Bom)** =  $r \geq 0,65$

## 4.2 Validade discriminante

O teste ou instrumento que se pretende acreditar deve apresentar correlações baixas com medidas que avaliam conceitos não relacionados ou ainda diferenciar grupos clínicos /especiais (e.g., um teste de inteligência deve distinguir claramente os desempenhos do grupo normativo comparativamente a um grupo com perturbação do desenvolvimento intelectual).



#### 4.2.1 Qualidade da(s) amostra(s)

##### a. Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### b. Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $N < 75$

**Pontuação 2 (Suficiente)** =  $75 \leq N < 150$

**Pontuação 3 (Bom)** =  $N \geq 150$

##### Observações:

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e o número de itens que o constituem, a população-alvo à qual o instrumento se destina, o constructo examinado, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

#### 4.2.2 Qualidade das medidas utilizadas

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### Regra orientadora:

**Pontuação 1 (Insuficiente)** = O(s) teste(s) seleccionado(s) para a avaliação da validade discriminante é/são inadequado(s) ou as suas propriedades psicométricas não são satisfatórias.

**Pontuação 2 (Suficiente)** = Adequada selecção do teste utilizado para a avaliação da validade discriminante e com propriedades psicométricas satisfatórias.

**Pontuação 3 (Bom)** = Adequada selecção de dois ou mais testes utilizados para a avaliação da validade discriminante e com boas propriedades psicométricas.

#### 4.2.3 Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### Regra orientadora:

**Pontuação 1 (Insuficiente)** = As correlações estimadas com amostras de dimensão adequada são estatisticamente significativas e o tamanho do efeito é médio ou elevado (i.e., verificam-se correlações positivas elevadas com um outro instrumento que mede constructo(s) independente(s) do constructo(s) examinado(s) pelo instrumento que está a ser acreditado).

**Pontuação 2 (Suficiente)** = As correlações estimadas com amostras de dimensão adequada, apesar de estatisticamente significativas, estão associadas a um baixo tamanho do efeito (i.e., verificam-se correlações reduzidas ou nulas com um outro instrumento que mede constructo(s) independente(s) do constructo(s) examinado(s) pelo instrumento que está a ser acreditado).

**Pontuação 3 (Bom)** = As correlações estimadas com amostras de dimensão adequada são próximas de 0 e com um tamanho do efeito estatisticamente não significativo (i.e., verificam-se correlações claramente reduzidas ou nulas com outros instrumentos que medem constructo(s) independente(s) do constructo(s) examinado(s) pelo instrumento que está a ser acreditado).

### 4.3 Parâmetros relativos à estrutura interna

Recolha de evidências de validade, baseada em parâmetros que analisam a estrutura interna do instrumento (i.e., intercorrelações, análise factorial e aplicação da TRI).

#### 4.3.1 Qualidade da(s) amostra(s)

##### a. Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### b. Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $< 5$  sujeitos por item

**Pontuação 2 (Suficiente)** =  $5 \leq$  sujeitos por item  $< 10$

**Pontuação 3 (Bom)** =  $\geq 10$  sujeitos por item

##### Observações:

- Em instrumentos com um reduzido número de itens, a dimensão da(s) amostra(s) deve ser superior a 100 sujeitos. Em qualquer caso deve ter-se em consideração a população-alvo à qual o instrumento se destina, a literatura e a investigação representativa realizada com outras adaptações do mesmo instrumento ou de instrumentos congéneres.

#### 4.3.2 Intercorrelações

##### a. Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

##### Observações:

- Valores de correlação elevados e significativos entre dimensões que, do ponto de vista teórico, estejam relacionadas entre si e correlações fracas entre dimensões não relacionadas.

### 4.3.3 Análise fatorial

#### a. Metodologia e qualidade dos resultados

0

SEM INFORMAÇÃO

1

INSUFICIENTE

2

SUFICIENTE

3

BOM

##### Regra orientadora:

**Pontuação 1 (Insuficiente)** = Os resultados não apoiam a estrutura do instrumento e/ou são baseados num estudo isolado de análise fatorial que apresenta limitações ao nível dos procedimentos de análise e/ou informação diminuta ao nível metodológico ou da interpretação dos resultados.

**Pontuação 2 (Suficiente)** = Os resultados apoiam a estrutura do instrumento, tanto no que se refere ao número de factores extraídos, como à sua interpretação, a partir de estudos de análise fatorial exploratória (AFE) ou análise fatorial confirmatória (AFC).

**Pontuação 3 (Bom)** = Os resultados apoiam a estrutura do instrumento, tanto no que se refere ao número de factores extraídos, como à sua interpretação, a partir de estudos de análise fatorial exploratória (AFE) e análise fatorial confirmatória (AFC), incluindo testes de invariância estrutural.

##### Observações:

- Ter em consideração a qualidade das decisões ao aplicar a Análise Fatorial Exploratória (AFE) e/ou Análise Fatorial Confirmatória (AFC), incluindo natureza e dimensão das amostras e software empregue. No caso da AFE: métodos de rotação, regras de decisão utilizadas na definição do número de factores (e.g., valores de saturação na definição dos factores, percentagem da variância explicada pelos factores). No caso da AFC: fundamentação teórica, especificação e identificação dos modelos testados e índices de ajustamento utilizados.

### 4.3.4 Aplicação da TRI

#### a. Metodologia e qualidade dos resultados

0

SEM INFORMAÇÃO

1

INSUFICIENTE

2

SUFICIENTE

3

BOM

##### Regra orientadora:

**Pontuação 1 (Insuficiente)** = Metodologia inadequada e/ou ausência de informação relativa ao contributo dos itens para uma estrutura interna adequada do instrumento.

**Pontuação 2 (Suficiente)** = Utilização da metodologia apropriada e informação que evidencie realização de estudos relativos ao enviesamento dos itens (Funcionamento Diferencial dos Itens - FDI) com um número limitado de grupos relevantes (e.g., grupos/variáveis sociodemográficas como idade e/ou sexo). Apresenta informação limitada sobre o contributo dos itens para uma estrutura interna adequada do instrumento.

**Pontuação 3 (Bom)** = Utilização da metodologia apropriada e informação relevante sobre diversos estudos acerca do enviesamento dos itens (FDI) realizados com vários grupos/variáveis relevantes (e.g., idade, sexo, língua materna, grupos especiais relevantes, clínicos, educativos, outros). Identifica o contributo dos itens para uma estrutura interna adequada do instrumento.

##### Observações:

- No Funcionamento Diferencial dos Itens (FDI) importa indicar modelos utilizados, considerando número e tipo de parâmetros relativos às pessoas, número e tipo de parâmetros relativos aos itens e a função matemática que relaciona parâmetros das pessoas e dos itens aos dados observados. Especificar análises realizadas relativas ao FDI. Assinalar número de itens do instrumento, natureza e dimensão das amostras e metodologia utilizada (Thomas, 2019).

### 4.4 Parâmetros relativos à comparação de grupos

Recolha de evidências de validade, baseada em parâmetros relativos à comparação de grupos (i.e., análise das diferenças entre grupos relevantes).

#### 4.4.1 Diferenças entre grupos relevantes

##### a. Qualidade dos resultados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = São definidas hipóteses de validação claras e adequadas, mas não são observadas diferenças significativas no sentido esperado.

**Pontuação 2 (Suficiente)** = São definidas hipóteses de validação claras e adequadas e são observadas diferenças significativas no sentido esperado.

**Pontuação 3 (Bom)** = São definidas hipóteses de validação claras e adequadas, são observadas diferenças significativas no sentido esperado e é prestada atenção ao tamanho do efeito.

**Observações:**

- Podem ser grupos naturais - por exemplo, grupos sociodemográficos ou experimentais.

#### Avaliação global da validade

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção. Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar:

**1\***

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

No caso de avaliação “Insuficiente”, justificar:

## F. NORMAS

Nesta secção são considerados dois pontos de avaliação distintos:

- Normas que constituem a base de uma interpretação normativa (i.e., deriva da comparação da pontuação da pessoa avaliada com a distribuição das pontuações observadas num grupo de referência);
- Normas que constituem a base de uma interpretação relativa a um critério ou domínio (neste caso, são requeridos pontos de corte estabelecidos através de diferentes critérios, como a avaliação de peritos ou dados provenientes de estudos empíricos, com vista a uma tomada de decisão clínica, educacional ou outra).

##### Adequação do método ao instrumento

Aplicável

Não aplicável (avançar para a secção G: “Sistema de Cotação e de Interpretação”)

## 1. Tipos de Normas | Assinalar todas as opções que se aplicam

Dados normativos (avaliar o ponto 2)

Critério(s) ou dados de referência (avaliar o ponto 3)

Outro(s) referencial/ais de interpretação

Indicar qual ou quais:

## 2. Dados normativos

### 2.1 Critérios de selecção da(s) amostra(s)

Aleatória

Estratificada

Sistemática

Cluster

Faseada

Quotas

Bola de neve

Orientada

Outro(s) tipo(s) de amostragem

Indicar qual ou quais:

### 2.2 Qualidade da(s) amostra(s)

#### 2.2.1 Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INADEQUADOS

**2**

ADEQUADOS

**Regra orientadora:**

**Pontuação 1 (Inadequados)** = Amostra(s) inadequada(s) e/ou não representativa(s) da população-alvo.

**Pontuação 2 (Adequados)** = Amostra(s) adequada(s) e representativa(s) da população-alvo.

**Observações:**

- O número e a natureza da(s) amostra(s) devem atender aos grupos específicos aos quais o instrumento se destina.
- Considerar a representatividade para o domínio de aplicação pretendido e a informação relativa a diferenças entre grupos relevantes.

## 2.2.2 Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

### Regra orientadora:

**Pontuação 1 (Insuficiente)** =  $N < 200$  para a amostra global

**Pontuação 2 (Suficiente)** =  $200 \leq N < 400$  para a amostra global

**Pontuação 3 (Bom)** =  $N \geq 400$  para a amostra global

Quando se trate de amostragem contínua, a dimensão da amostra pode ser inferior em cada subgrupo e o número de subgrupos deverá ser representativo do constructo avaliado.

**Pontuação 1 (Insuficiente)** =  $N < 50$  em cada subgrupo.

**Pontuação 2 (Suficiente)** =  $50 \leq N < 100$  em cada subgrupo.

**Pontuação 3 (Bom)** =  $N \geq 100$  em cada subgrupo.

### Observações:

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e a população-alvo à qual o instrumento se destina. A dimensão de amostras de grupos específicos deve ter em consideração a frequência ou a prevalência na população-alvo (e.g., grupos clínicos, faixas etárias, grupos minoritários).
- A avaliação da dimensão da(s) amostra(s) deve ter em consideração publicações de referência sobre o instrumento.
- Caso se trate da adaptação de um instrumento, só deverão ser consideradas as normas internacionais se houver evidências de equivalência entre as duas versões do teste aplicado nas respectivas populações-alvo. Nesse caso, tais evidências devem ser sumariamente apresentadas e referenciadas.

## 2.3 Actualidade dos dados normativos

**0**

SEM INFORMAÇÃO

**1**

INADEQUADOS

**2**

ADEQUADOS

### Regra orientadora:

**Pontuação 1 (Insuficiente)** = Superior a 20 anos

**Pontuação 2 (Suficiente)** = Entre 10 e 20 anos

**Pontuação 3 (Bom)** = Inferior a 10 anos

### Observações:

- Considerar a actualidade das normas, tendo em consideração o intervalo temporal entre a recolha da amostra e a publicação dos dados normativos de referência.

## 2.4 Estatística descritiva dos resultados obtidos pelos grupos normativos

**0**

SEM INFORMAÇÃO

**1**

INADEQUADOS

**2**

ADEQUADOS

### Regra orientadora:

**Pontuação 1 (Inadequada)** = Informação incompleta e/ou pouco pormenorizada sobre a estatística descritiva dos resultados obtidos pelos grupos normativos.

**Pontuação 2 (Adequada)** = Informação completa e pormenorizada sobre a estatística descritiva dos resultados obtidos pelos grupos normativos.

## Avaliação global dos dados normativos

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção. Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar:

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Integração e ponderação em função das pontuações atribuídas às alíneas do ponto 2.

## 3. Critério(s) e/ou dados de referência

### 3.1 Qualidade da(s) amostra(s)

#### 3.1.1 Número e natureza da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INADEQUADOS

**2**

ADEQUADOS

**Regra orientadora:**

**Pontuação 1 (Inadequados)** = Amostra(s) inadequada(s) e/ou não representativa(s) do critério

**Pontuação 2 (Adequados)** = Amostra(s) adequada(s) e representativa(s) do critério

#### 3.1.2 Dimensão da(s) amostra(s)

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** =  $N < 40$

**Pontuação 2 (Suficiente)** =  $40 \leq N < 80$

**Pontuação 3 (Bom)** =  $N \geq 80$

**Observações:**

- A dimensão da(s) amostra(s) deve ter em consideração o tipo de instrumento e a frequência ou a prevalência na população-alvo à qual o instrumento se destina.
- A avaliação da dimensão da(s) amostra(s) deve ter em consideração publicações de referência sobre o instrumento.

### 3.2 Pontos de corte | Assinalar todos os critérios utilizados para a definição dos pontos de corte

Acordo entre peritos (avaliar a alínea 3.2.1)

Critérios de natureza empírica (avaliar a alínea 3.2.2)

### 3.2.1 Acordo entre peritos

#### a. Número de peritos e metodologia utilizada

**0**  
SEM INFORMAÇÃO

**1**  
INADEQUADOS

**2**  
ADEQUADOS

**Regra orientadora:**

**Pontuação 1 (Inadequados)** < 3 peritos e/ou metodologia inadequada

**Pontuação 2 (Adequados)** ≥ 3 peritos e metodologia adequada

**Observações:**

- A avaliação deste ponto deve considerar a adequação da(s) metodologia(s) utilizada(s) na definição dos pontos de corte e as referências existentes na literatura especializada.

#### b. Indicador(es) de acordo entre peritos | Assinalar todas as opções aplicáveis

Coefficiente Kappa

Coefficiente de Correlação Intraclasse (ICC)

Outro(s)

Indicar qual ou quais:

#### c. Grau de acordo entre peritos

**0**  
SEM INFORMAÇÃO

**1**  
INSUFICIENTE

**2**  
SUFICIENTE

**3**  
BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = Kappa e/ou ICC < 0,60

**Pontuação 2 (Suficiente)** =  $0,60 \leq$  Kappa e/ou ICC < 0,70

**Pontuação 3 (Bom)** = Kappa e/ou ICC ≥ 0,70

**Observações:**

- Não existem valores de referência inequívocos para os coeficientes Kappa ou Intraclasse. Há, contudo, algum consenso quanto à inadequabilidade de valores inferiores a 0,60.

### 3.2.2 Critérios de natureza empírica

#### a. Estudos realizados e indicadores utilizados

**0**  
SEM INFORMAÇÃO

**1**  
INADEQUADOS

**2**  
ADEQUADOS

**Observações:**

- A avaliação deste ponto deve ter em consideração a robustez dos estudos, o rigor das metodologias e a adequação dos indicadores.



### 3.2.3 Actualidade dos estudos

#### a. Estudos realizados e indicadores utilizados

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 1 (Insuficiente)** = Superior a 20 anos

**Pontuação 2 (Suficiente)** = Entre 10 e 20 anos

**Pontuação 3 (Bom)** = Inferior a 10 anos

**Observações:**

- Ter em consideração o intervalo temporal entre a recolha da amostra e a publicação dos critérios e/ou dados de referência.

#### Avaliação global do(s) critério(s) e/ou dados de referência

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção. Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar:

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Integração e ponderação em função das pontuações atribuídas às alíneas do ponto 3.

#### Avaliação global das normas

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção.

Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar em risco uma adequada utilização do instrumento.

**1\***

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

\*No caso de avaliação "Insuficiente", justificar:

# G. SISTEMA DE COTAÇÃO E DE INTERPRETAÇÃO

## 1. Sistema de cotação

### 1.1 Procedimentos de cotação | Assinalar todas as opções que se aplicam

#### Procedimentos manuais de registo e/ou cotação

Procedimentos manuais simples (apenas envolve rotinas administrativas)

Procedimentos manuais complexos (envolve formação específica)

#### Procedimentos automatizados de registo de respostas com base:

Na introdução das respostas, pelo avaliador

Nas folhas de resposta, com sistema próprio integrado (p.e., chave de cotação integrada na própria folha)

Na leitura ótica das folhas de resposta

#### Procedimentos automatizados de cotação e obtenção de resultados, por recurso a:

Software da entidade que comercializa o instrumento

Serviço prestado pela empresa que comercializa o instrumento (folhas de resposta enviadas à empresa)folha)

Outro(s) procedimentos

Indicar qual ou quais:

### 1.2 Instruções de cotação | Aplica-se apenas a instrumentos que não recorrem a software de cotação

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Instruções de cotação explícitas, claras e pormenorizadas, garantindo uniformização e objectividade. No caso dos instrumentos que envolvem cotação ou codificação, a partir da análise qualitativa das respostas, indicação de critérios e procedimentos explícitos de análise e de posterior cotação/codificação. A pontuação 3 exige a existência de exemplos de critérios de análise e exemplos de resposta para cada nível, tipo ou categoria de cotação/codificação.

## Avaliação global do sistema de cotação

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção. Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar:

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Integração e ponderação em função das pontuações atribuídas às alíneas do ponto 1.

## 2. Resultados finais | Este ponto refere-se aos resultados que são objecto de interpretação

### 2.1 Tipo(s) de resultados | Assinalar todas as opções que se aplicam

Apenas resultados brutos

Resultados transformados:

Com base em tabelas de normas (resultados normalizados)

Transformação linear (resultados não normalizados)folha)

Resultados expressos em percentis

Resultados padronizados:

T (M = 50; DP = 10)

QI ou equivalente (M = 100; DP = 15)

Z (M = 0; DP = 1)

Outro tipo de distribuição

Indicar qual ou quais:

Resultados expressos em outra escala de medida:

Indicar qual ou quais:

Resultados não quantitativos:

Indicar as categorias em que os dados qualitativos são apresentados e a secção do manual em que as mesmas são explicitadas:

## Avaliação global dos resultados finais

Integração e ponderação em função das pontuações atribuídas nos diversos pontos e alíneas desta secção. Esta avaliação deverá ser classificada como insuficiente, sempre que uma das alíneas avaliadas colocar:

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

### Regra orientadora:

Tipo de resultados e de distribuição adequados à natureza e objectivos do instrumento.

## 3. Sistema de interpretação

### 3.1 Interpretação referenciada à norma

Aplicável aos instrumentos em que a interpretação (referenciada à norma) decorre da comparação dos resultados do desempenho da pessoa avaliada com a distribuição de resultados no grupo de referência correspondente, ou seja, instrumentos concebidos para serem utilizados com recurso a normas.

#### Aplicabilidade ao instrumento

Aplicável

Não aplicável (avançar para a alínea 3.2)

Assinalar a opção que melhor descreve o tipo de interpretação

Interpretação de tipo descritivo e de baixa complexidade (i.e., indica o nível de desempenho ou de competência ou de traço/estado tendo por referência as normas)

Interpretação de média complexidade (i.e., descreve e relaciona diferentes dimensões analisadas tendo por referência as normas)

Interpretação de elevada complexidade (i.e., descreve padrões ou perfis de resultados, e analisa os efeitos das interações entre dimensões avaliadas)

### 3.2 Interpretação referenciada a um critério

Aplicável aos instrumentos em que a interpretação (referenciada a um critério ou domínio) tem por referência pontos de corte que indicam a probabilidade de (in)existência de determinadas características ou condição (e.g., competências, sintomas, diagnóstico clínico), i.e., instrumentos concebidos para serem utilizados sem recurso a normas. Os pontos de corte podem ser estabelecidos a partir de dados empíricos ou por consenso de especialistas.

#### Aplicabilidade ao instrumento

Aplicável

Não aplicável (avançar para a alínea 3.3)

Assinalar a opção que melhor descreve o tipo de interpretação

Interpretação de tipo descritivo/avaliativo e de baixo nível inferencial (i.e., interpretação baseada em dados empíricos obtidos a partir do instrumento ou a ele associados)

Interpretação de média complexidade (i.e., interpretação baseada em procedimentos analíticos relacionados com o constructo e os fundamentos teórico do instrumento que envolve um moderado nível inferencial)

Interpretação de elevada complexidade (i.e., interpretação que exige procedimentos analíticos relacionados com o constructo e os fundamentos teóricos do instrumento e a articulação com outros conceitos e teorias psicológicas/psicopatológicas, envolvendo um elevado nível inferencial.

### 3.3 Outro(s) tipo(s) de interpretação

Indicar qual ou quais:

### 3.4 Orientações para interpretação

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

**Pontuação 3 (Bom)** = Orientações gerais claras, objectivas e explícitas que permitam uma análise e interpretação de resultados consistente e uniforme (i.e., sem significativa variabilidade quando seguidas por diferentes avaliadores).

São também identificados:

a) procedimentos de controlo da qualidade de resposta (e.g., invariabilidade, omissão ou duplicação de resposta);

b) critérios ou condições que limitem a análise e a interpretação global dos resultados (e.g., número máximo de omissões de resposta ou número mínimo de respostas aceitáveis para interpretação do protocolo; comportamentos que denunciem incompreensão ou reduzido empenhamento nas tarefas sob avaliação).

### Avaliação global do sistema de interpretação

**0**

SEM INFORMAÇÃO

**1**

INSUFICIENTE

**2**

SUFICIENTE

**3**

BOM

**Regra orientadora:**

Integração e ponderação em função das pontuações atribuídas às diversas alíneas do ponto 3

### 4. Relatórios gerados automaticamente

Consideram-se nesta categoria os documentos gerados automaticamente por software específico, em que são apresentados resultados associados a uma descrição das dimensões avaliadas e da posição relativa da pessoa avaliada face a uma dada norma ou critério. Nesse sentido, este ponto apenas é avaliado caso o instrumento disponha desta opção.

**Aplicabilidade ao instrumento**

Aplicável

Não aplicável (avançar para a Avaliação do sistema de cotação e de interpretação)

#### 4.1 Tipo de relatório | Assinalar as opções que se aplicam

Relatório não-técnico (dirigido à pessoa avaliada)

Procedimentos manuais complexos (envolve formação específica)

Relatório não-técnico (dirigido a Terceiros)

#### 4.2 Suscetibilidade de modificação | Assinalar a opção que se aplica

Não permite modificações

Permite modificações limitadas (e.g., dados biográficos)

Permite amplas modificações (i.e., é um documento editável)

#### 4.3 Nível de complexidade

BAIXO

MÉDIO

ELEVADO

**Regra orientadora:**

**Baixa complexidade:** descreve as dimensões avaliadas e os resultados ao nível das medidas finais, parciais e/ou globais;

**Média complexidade:** descreve e relaciona algumas das dimensões avaliadas;

**Elevada complexidade:** descreve padrões e configurações ou perfis de resultados, e analisa os efeitos das interações entre as dimensões avaliadas.

#### Avaliação global dos relatórios gerados automaticamente

0

SEM INFORMAÇÃO

1

INSUFICIENTE

2

SUFICIENTE

3

BOM

**Regra orientadora:**

Atender à adequação do tipo de relatório ao destinatário, à clareza da linguagem, das análises, das interpretações e das conclusões, bem como à qualidade dos fundamentos e consistência dos conteúdos e à sua potencial utilidade.

**Avaliação global do sistema de cotação e de interpretação**

**1\***  
INSUFICIENTE

**2**  
SUFICIENTE

**3**  
BOM

\*No caso de avaliação "Insuficiente", justificar:

**VOLTAR AO ÍNDICE**

# III. PARECER FINAL

Sim

Não\*

preencher também o campo comentários

O instrumento de avaliação cumpre, globalmente, com os requisitos técnico-científicos para a sua utilização no âmbito da prática da Avaliação Psicológica?

\* Um instrumento cuja apreciação global é a de que não cumpre com os requisitos técnico-científicos para a sua utilização no âmbito da prática da Avaliação Psicológica apenas poderá ser utilizado para efeitos de investigação, tendo em vista a sua melhoria.

## Comentários

Uma apreciação global negativa terá de ser fundamentada com a indicação dos requisitos mínimos não cumpridos, através da apresentação de resumo das classificações/observações que possam guiar as melhorias a desenvolver.

Sugestões adicionais são aconselháveis sempre que se considere que há uma margem de melhoria substancial do instrumento ou das condições da sua utilização.

VOLTAR AO ÍNDICE



# IV. REFERÊNCIAS BIBLIOGRÁFICAS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row.

European Federation of Psychologists Associations, Task Force of the Board of Assessment of EFPA. (2013). EFPA Review Model for the description and evaluation of psychological instruments: Test Review Form and Notes for Reviewers (Version 4.2.6). Retrieved from <http://assessment.efpa.eu/download/650d0d4ecd407a51139ca44ee704fda4>

Hernández, A., Ponsoda, V., Muñiz, J., Prieto, G., & Elosua, P. (2016). Revisión del modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 37(3), 192-197.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>

Lord, F.M. (1980). Applications of Item Response Theory To Practical Testing Problems (1st ed.). Routledge. <https://doi.org/10.4324/9780203056615>

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

Parshall, C. G., Spray, J. A., Kalohn, J., & Davey, T. (2002). Practical Considerations in Computer-based Testing. New York: Springer Verlag.

Reise, S. P. & Haviland, M.G. (2005). Item Response Theory and the Measurement of Clinical Change. *Journal of Personality Assessment*, 84(3), 228-238. doi:10.1207/s15327752jpa8403\_02

Schneider, R. J., & Hough, L. M. (1995). Personality and industrial/organizational psychology. In C. L. Cooper &

I. T. Robertson (Eds.), *International review of industrial and organizational psychology*, (Vol. 10; pp. 75-130). Chichester, UK: John Wiley & Sons, Inc.

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment*, 31(12), 1442-1455. doi:10.1037/pas0000597

**VOLTAR AO ÍNDICE**

