



## **COMISIA METODOLOGICĂ**

### **PROIECT DE MODIFICARE A NORMELOR DE AVIZARE A METODELOR ȘI TEHNICILOR DE EVALUARE ȘI ASISTENȚĂ PSIHOLOGICĂ**

**Document elaborat în conformitate cu standardele actuale EFPA, Board of Assessment -  
EFPA REVIEW MODEL FOR THE DESCRIPTION AND EVALUATION OF  
PSYCHOLOGICAL AND EDUCATIONAL TESTS, versiunea 4.2.6. Autori: Arne Evers,  
Carmen Hagemester, Andreas Høstmælingen , Patricia Lindley, José Muñiz, Anders  
Sjöberg.**

**Proiect propus de Comisia Metodologică, în baza obiectivelor electorale și liniilor  
directoare.**

## Cuprins

### Introducere

#### Partea întâi – Descrierea instrumentului

1. Descrierea generală
2. Clasificare
3. Măsurare și scorare
4. Rapoarte generate pe calculator
5. Condiții și costuri de livrare

#### Partea a doua – Evaluarea instrumentului

6. Calitatea justificării utilității instrumentului, a modalității de prezentare și a informațiilor oferite
  - 6.1. Calitatea explicării utilității instrumentului
  - 6.2. Adecvarea documentației disponibile utilizatorilor (manuale de utilizare și tehnice, etaloanele etc.)
  - 6.3. Calitatea instrucțiunilor procedurale oferite utilizatorilor
7. Calitatea materialelor de testare
  - 7.1. Calitatea materialelor aferente testelor creion-hârtie
  - 7.2. Calitatea materialelor aferente testării computerizate (TC) și a celei web (TW)
8. Etaloanele
  - 8.1. Interpretarea bazată pe norme
  - 8.2. Interpretare bazată pe criteriu
9. Fidelitatea
10. Validitatea
  - 10.1. Validitatea de construct
  - 10.2. Validitatea relativă la criteriu
  - 10.3. Validitatea generală
11. Calitatea rapoartelor generate de computer
12. Evaluarea finală și tipuri de rezoluții

#### Partea a treia – Avizarea metodelor de asistență / intervenție psihologică

#### Dispoziții finale

#### Referințe bibliografice

**Anexă** - Ghid al punctelor critice în cazul unui instrument tradus și / sau adaptat dintr-un context non-local

## Introducere

Scopul principal al **Proiectului de modificare a normelor de avizare a metodelor și tehnicilor de evaluare și asistență psihologică** este de a oferi o descriere și o evaluare detaliată și riguroasă a testelor, scalelor și chestionarelor de evaluare psihologică utilizate în psihologia aplicată în domeniile muncii, educației, sănătății și în alte contexte. Acest model va fi disponibil utilizatorilor de teste și profesioniștilor, în scopul optimizării activitatea de testare psihologică și oferirii de suport în a lua deciziile de evaluare corecte. Proiectul este fundamentat pe standardul EFPA în domeniu. Urmând Standardele pentru Testarea Educațională și Psihologică, eticheta de test este utilizată pentru orice "... instrument evaluativ sau procedură în care un eșantion al comportamentului examinatului într-un domeniu specific este obținut și apoi evaluat și scorat utilizând un proces standardizat" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 3). Astfel, acest model de evaluare se aplică tuturor instrumentelor acoperite sub această definiție, indiferent dacă sunt denumite scale, chestionare, tehnici proiective, sau în alt mod.

Versiunea originală a modelului de evaluare a testelor EFPA a fost bazată pe o serie de surse, incluzând: BPS Test Review Evaluation Form (dezvoltat de Newland Park Associates Limited, NPAL, iar apoi adoptat de BPS Steering Committee on Test Standards); chestionarul spaniol pentru evaluarea testelor psihometrice (dezvoltat de Asociația Spaniolă de Psihologie) și sistemul de evaluare a calității testului (elaborat de Comitetul Olandez de Teste și Testare al Asociației Psihologilor din Olanda). O mare parte din conținut a fost adaptat cu permisiunea proformei de revizuire elaborată inițial în 1989 de către Newland Park Associates Ltd., pentru o analiză a testelor utilizate de agenții de formare în Marea Britanie (vezi Bartram, Lindley & Foster, 1990). Aceasta a fost utilizată și dezvoltată ulterior pentru o serie de revizuri BPS ale instrumentelor utilizate în evaluarea ocupațională (de exemplu, Bartram, Lindley, & Foster, 1992; Lindley et al., 2001). Prima versiune a modelului de evaluare EFPA a fost compilată și editată de Dave Bartram (Bartram, 2002a, 2002b) în urma unui atelier inițial EFPA din martie 2000 și a următoarelor runde de consultare. O actualizare majoră și o revizuire a fost efectuată de Patricia Lindley, Dave Bartram și Natalie Kennedy pentru utilizarea în sistemul de revizuire BPS (Lindley et al, 2004). Acest model a fost adoptat ulterior de EFPA în 2005 (Lindley et al., 2005), cu revizuri minore în 2008 (Lindley et al., 2008). Versiunea actuală a modelului a fost pregătită de către un grup operativ al Comitetului de evaluare EFPA, al cărui membri sunt Arne Evers (președinte, Olanda), Carmen Hagemester (Germania), Andreas Høstmælingen (Norvegia), Patricia Lindley (Marea Britanie) José Muñiz (Spania) și Anders Sjöberg (Suedia). În această versiune au fost integrate notele și lista de verificări pentru testele traduse și adaptate produse de Pat Lindley și Editorii Consultanți de Recenzare a Testelor din UK (Lindley, 2009). Textele unor pasaje majore actualizate se bazează pe sistemul revizuit de evaluare al Olandei pentru calitatea testului (Evers, Lucassen, Meijer, & Sijtsma, 2010; Evers, Sijtsma, Lucassen, & Meijer, 2010).

**Proiectul de modificare a normelor de avizare a metodelor și tehnicilor de evaluare și asistență psihologică** este structurat în două părți principale. În prima parte (Descrierea instrumentului), toate caracteristicile testului evaluat sunt descrise în detaliu. În a doua parte

(Evaluarea instrumentului), sunt evaluate proprietățile fundamentale ale testului: materiale de testare, norme, fidelitate, validitate și rapoarte generate de calculator, inclusiv o evaluare finală globală.

Acest model este destinat utilizării de către doi evaluatori independenți, într-un proces de evaluare peer review, similar cu evaluarea obișnuită a lucrărilor științifice și a proiectelor. Președintele Comisiei Metodologice va superviza evaluările și poate apela la un al treilea evaluator, dacă se vor găsi discrepanțe semnificative între cele două evaluări. Se recomandă ca evaluările să fie îndreptate către practicienii calificați, deși ar trebui să fie de interes și pentru cadre universitare, autori de testare și specialiști în psihometrie și teste psihologice. O altă problemă-cheie este publicarea rezultatelor evaluării unui test. Rezultatele trebuie să fie disponibile pentru toți profesioniștii și utilizatorii, pe site-ul Colegiului Psihologilor din România, și ar putea fi publicate și de terțe părți sau prin intermediul altor canale de mass-media, cum ar fi reviste de specialitate sau cărți.

În cazul metodelor de asistență psihologică, evaluarea se realizează prin analiza dovezilor științifice invocate.

Obiectivul fundamental al elaborării acestui document îl constituie armonizarea procedurilor și criteriilor de analiză și evaluare a metodelor și tehnicilor de evaluare și asistență psihologică, în conformitate cu standardele europene în domeniu.

## PARTEA ÎNTÂI – DESCRIEREA INSTRUMENTULUI

### 1. Descrierea generală

Această secțiune are scopul menirea de a oferi informații de bază, utile pentru identificarea instrumentului și a modalității prin care acesta poate fi obținut. Trebuie să conțină titlul instrumentului, editorul/editura sau distribuitorul, autorii, data primei publicări și data versiunii supuse evaluării.

Întrebările 1.1.1 - 1.7.3 trebuie să fie formulate clar. Ele fac referire la informații factice, însă pentru a completa câmpurile despre domeniile de conținut este nevoie de anumite evaluări.

	<b>Evaluator<sup>1</sup></b>	
	<b>Data evaluării</b>	
	<b>Data evaluării anterioare (dacă se aplică)<sup>2</sup></b>	
1.1.1.	<b>Denumirea testului (versiunea locală)</b>	
1.1.2.	<b>Denumirea prescurtată a testului (dacă se aplică)</b>	
1.2.	<b>Denumirea originală a testului (dacă versiunea locală este o adaptare)</b>	
1.3.	<b>Autorii testului original</b>	
1.4.	<b>Autorii adaptării locale</b>	
1.5.	<b>Distribuitorul/editorul local</b>	
1.6.	<b>Editura unde a fost publicată versiunea originală a testului (dacă aceasta diferă de cea actuală sau de distribuitorul actual)</b>	
1.7.1.	<b>Data publicării actualei revizuirii sau ediții</b>	
1.7.2.	<b>Data publicării versiunii adaptate pentru uz local</b>	
1.7.3.	<b>Data publicării testului original</b>	

<sup>1</sup> Numele evaluatorului poate să nu fie făcut public.

<sup>2</sup> Această informație este furnizată de editorul testului.

**Descrierea generală a instrumentului** Scurtă descriere non-evaluativă, cuprinzătoare (200-600 de cuvinte)

În acest câmp trebuie oferită o descriere concisă, non-evaluativă a instrumentului. Descrierea trebuie să confere cititorului o idee clară despre elementele definiției ale instrumentului – ce conține, scalele din care se compune, ce variabile pretinde că măsoară etc. Formularea ar trebui să fie cât mai neutră posibil. Ar trebui menționat și scopul pentru care a fost conceput instrumentul, accesibilitatea și caracteristicile grupurilor pe care poate fi aplicat instrumentul, etaloanele, dar și alte aspecte de interes, caracteristici speciale ale instrumentului sau orice informații relevante din istoricul dezvoltării instrumentului. Această descriere poate fi succintă (200-300 de cuvinte). Totuși, pentru unele instrumente mai complexe, formate din mai multe scale, e nevoie ca descrierea să fie mai lungă (300-600 de cuvinte). Textul ar trebui scris în așa fel încât să poată constitui o descriere de sine stătătoare a instrumentului. Descrierea ar trebui să evidențieze toate versiunile disponibile ale instrumentului, despre care au fost făcute referiri și în alte pagini ale acestui formular.

Răspunsul la acest item trebuie formulat pe baza informațiilor oferite de autor/editor, iar acuratețea sa trebuie verificată de către evaluator.

## 2. Clasificare

<p>2.1.</p>	<p><b>Domenii de conținut</b> (<i>selectați toate variantele care se aplică</i>)</p> <p>Trebuie identificate domeniile de conținut specificate de autor/editor. Dacă acestea nu sunt clare, acest lucru trebuie menționat, iar răspunsurile cele mai potrivite pentru itemii din secțiunea 2.1 trebuie estimate pe baza informațiilor regăsite în manualul instrumentului (eșantioanele folosite pentru standardizare, aplicații, validare etc.).</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Aptitudine/Abilitate - Generală</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Abilități manuale/dexteritate</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Mecanică</li> <li><input type="checkbox"/> Aptitudine/Abilitate - De învățare/memorie</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Non-verbală/abstractă/inductivă</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Numerică</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Viteză perceptivă/de verificare</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Senzorio-motorie</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Spațială/vizuală</li> <li><input type="checkbox"/> Aptitudine/Abilitate - Verbală</li> <li><input type="checkbox"/> Atenție/concentrare</li> <li><input type="checkbox"/> Credințe</li> <li><input type="checkbox"/> Stiluri cognitive</li> <li><input type="checkbox"/> Tulburare și patologie</li> <li><input type="checkbox"/> Funcționare familială</li> <li><input type="checkbox"/> Funcționare în grup</li> <li><input type="checkbox"/> Interese</li> <li><input type="checkbox"/> Motivație</li> <li><input type="checkbox"/> Funcționare organizațională, măsurători agregate, climat etc.</li> <li><input type="checkbox"/> Personalitate – Trăsătură</li> <li><input type="checkbox"/> Personalitate – Tip</li> <li><input type="checkbox"/> Personalitate – Stare</li> <li><input type="checkbox"/> Calitatea vieții</li> <li><input type="checkbox"/> Performanță școlară/academică (test educațional)</li> <li><input type="checkbox"/> Funcționare școlară sau educațională</li> <li><input type="checkbox"/> Judecată situațională</li> <li><input type="checkbox"/> Stres/burnout</li> <li><input type="checkbox"/> Rezultate terapeutice</li> <li><input type="checkbox"/> Valori</li> <li><input type="checkbox"/> Stare de bine</li> <li><input type="checkbox"/> Altele (vă rugăm să descrieți):</li> </ul>
<p>2.2.</p>	<p><b>Arii de aplicabilitate principale sau dorite</b> (<i>vă rugăm să le selectați pe acelea care se aplică</i>)</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Clinică</li> <li><input type="checkbox"/> Consiliere și orientare în alegerea carierei</li> <li><input type="checkbox"/> Domeniul judiciar</li> <li><input type="checkbox"/> Sănătate generală, viață și starea de bine</li> </ul>

	<p>Trebuie identificate ariile de aplicabilitate pentru care a fost conceput instrumentul, specificate de autor/editor. Dacă acestea nu sunt clare, acest lucru trebuie menționat, iar răspunsurile cele mai potrivite pentru itemii din secțiunea 2.2 trebuie estimate pe baza informațiilor regăsite în manualul instrumentului (eșantioanele folosite pentru standardizare, aplicații, validare etc.).</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Domeniul neurologic</li> <li><input type="checkbox"/> Sport și timp liber</li> <li><input type="checkbox"/> Muncă și ocupație</li> <li><input type="checkbox"/> Altele (vă rugăm să descrieți):</li> </ul>
2.3.	<p><b>Descrierea populațiilor pentru care a fost conceput testul</b></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor. Pentru unele teste, aceste informații pot fi foarte generale (de ex., „adulți”). Pentru altele pot fi mai specifice (de ex., „muncitorii care lucrează manual”, sau „băieți cu vârsta cuprinsă între 10 și 14 ani”). Doar populațiile menționate pot fi notate aici. Dacă acest lucru pare nepotrivit, trebuie adăugat un comentariu în acest sens la secțiunea „Evaluare” a acestui formular.</p>	
2.4.	<p><b>Numărul scalelor și o scurtă descriere a variabilei/variabilelor măsurate cu ajutorul instrumentului</b></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor. Vă rugăm să precizați numărul de scale (dacă sunt cel puțin două) și să oferiți o scurtă descriere a fiecărei scale, dacă denumirile nu sunt auto-explicative.</p> <p>Prezentarea instrumentului ar trebui să includă și o discuție despre alte scoruri derivate, dacă aceste scoruri sunt folosite frecvent și dacă sunt descrise în documentația standard – de ex., scoruri pentru trăsături primare și scoruri pentru trăsături Big Five secundare</p>	



	<p>pentru un instrument multidimensional de evaluare a personalității; sau un subtest, factor sau scoruri totale la un test de inteligență.</p>	
2.5.	<p><b>Modalitatea de răspuns</b></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor. Orice alt echipament necesar, diferit de ceea ce apare în lista de opțiuni (de ex., reportofon) trebuie descris aici. De asemenea, orice condiții speciale de testare trebuie descrise. Se presupune că există „condiții standard de testare” pentru o evaluare supravegheată/supervizată. Acestea ar include liniște, o cameră bine iluminată și bine ventilată, precum și un spațiu adecvat dotat cu birouri și scaune pentru evaluator/i și candidat/candidați.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Interview oral</li> <li><input type="checkbox"/> Creion-hârtie</li> <li><input type="checkbox"/> Operațiuni manuale (fizice)</li> <li><input type="checkbox"/> Observație directă</li> <li><input type="checkbox"/> Calculator</li> <li><input type="checkbox"/> Altele (indicați):</li> </ul>
2.6.	<p><b>Criterii pentru persoana căreia i se aplică testul</b></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor. Ce capacități și abilități e necesar ca persoana evaluată să dețină pentru ca aceasta să efectueze testul așa cum s-a intenționat și pentru ca interpretarea rezultatului la test să poată fi făcută corect? De regulă, este clar dacă absența unor prerechizite face imposibilă completarea testului (de ex., o persoană nevăzătoare nu poate primi un test creion-hârtie tipic), însă cerințele necesare trebuie clasificate după cum urmează:</p> <ul style="list-style-type: none"> <li>• „Irelevant/ne-necesar”, înseamnă că această abilitate nu este necesară deloc (de ex., capacități manuale pentru a răspunde verbal la întrebări orale).</li> </ul>	<p>Capacități manuale (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul> <p>Neîndemânare (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul> <p>Vedere (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul> <p>Auz (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul> <p>Comenzile date de limbajul testului (înțelegerea vorbirii) (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> </ul>

	<ul style="list-style-type: none"> <li>• „Informații necesare oferite”, înseamnă că se menționează limitările posibile.</li> <li>• „Informații lipsă”, înseamnă că ar putea fi unele limitări pentru utilizatorii care nu dețin o anumită capacitate sau abilitate (așa cum se cunoaște din teorie sau pe baza rezultatelor empirice), dar că acest lucru nu rezultă clar din informațiile oferite de autor/editor (de ex., dacă testul este conceput într-o limbă diferită de limba natală a persoanei testate).</li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul> <p>Citit (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul> <p>Scris (<i>selectați una</i>)</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> irelevant / ne-necesar</li> <li><input type="checkbox"/> informații necesare oferite</li> <li><input type="checkbox"/> informații lipsă</li> </ul>
2.7.	<p><b>Formatul itemilor</b> (<i>selectați una</i>)</p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor.</p> <p>Se diferențiază două tipuri de răspunsuri multiple. Primul tip este valabil pentru testele în care respondentul trebuie să aleagă răspunsul corect din mai multe alternative, așa cum se întâmplă în cazul testelor de abilități (de ex., un test de raționament bazat pe imagini). Al doilea tip se referă la chestionarele în care nu există clar un răspuns corect. Acest format le solicită respondenților să facă alegeri pe baza unor seturi de doi sau mai mulți itemi proveniți din scale diferite (de ex., scalele dintr-un inventar de interese vocaționale sau dintr-un chestionar de personalitate). Acest format mai este numit și „multidimensional”, pentru că alternativele aparțin unor scale sau dimensiuni diferite. În acest caz este posibil ca răspunsurile să necesite o ierarhizare sau să fie nevoie de o selecție a unor opțiuni de tipul „mă reprezintă cel mai mult” – „mă reprezintă cel mai puțin”. Acest format poate să ducă la scale ipsative (vezi întrebarea 2.8).</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Răspuns multiplu (testarea abilităților, sau corect/greșit) Numărul alternativelor: ....</li> <li><input type="checkbox"/> Răspuns multiplu (alternative tip scală mixtă) Numărul alternativelor: ....</li> <li><input type="checkbox"/> Scală tip Likert Numărul alternativelor: ....</li> <li><input type="checkbox"/> Răspuns deschis</li> <li><input type="checkbox"/> Altele (vă rugăm să descrieți)</li> </ul>

	<p>Cotările pe o scală Likert presupun de asemenea ca persoana evaluată să aleagă dintr-un număr de alternative, însă diferența esențială față de formatul răspunsuri multiple este că scalele utilizate sunt unidimensionale (de ex., variază între „niciodată” și „întotdeauna” sau între „foarte improbabil” și „foarte probabil”), iar persoana evaluată nu trebuie să aleagă între alternative provenite din diferite dimensiuni. O scală este considerată tip Likert și dacă presupune doar două alternative pe o dimensiune (de ex., da/nu sau întotdeauna/niciodată).</p>	
2.8.	<p><b>Caracter ipsativ</b></p> <p>Așa cum a fost menționat la secțiunea 2.7, alternativele corespunzătoare formatului răspuns multiplu scale mixte pot duce la scoruri ipsative. Specificul scorurilor ipsative este că scorul la fiecare scală sau dimensiune depinde de scorurile la alte scale sau dimensiuni. În cazul instrumentelor ipsative în totalitate, suma scorurilor la fiecare scală este constantă pentru fiecare persoană. Alte proceduri de scorare pot determina caracterul ipsativ (de ex., scăderea mediei generale obținute de o persoană, din scorurile sale pentru fiecare scală).</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Da, răspunsuri multiple tip scale mixte ce duc la scoruri ipsative parțial sau total.</li> <li><input type="checkbox"/> Da, alte tipuri de itemi, ce presupun proceduri de scorare ce duc la scoruri ipsative parțial sau total.</li> <li><input type="checkbox"/> Nu, răspunsuri multiple tip scale mixte care NU duc la scoruri ipsative</li> <li><input type="checkbox"/> Nu este relevant</li> </ul>
2.9.	<p><b>Numărul total de itemi și numărul itemilor pe scală sau subtest</b></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor. Dacă instrumentul are mai multe scale sau subteste, indicați numărul total al itemilor și numărul itemilor corespunzători fiecărei scale sau subtest. Dacă există itemi care fac parte</p>	

	din compoziția mai multor scale, acest detaliu trebuie documentat.	
2.10.	<p><b>Modul de utilizare recomandat (condițiile în care instrumentul a fost dezvoltat și validat) (selecțiați tot ce se aplică)</b></p> <p>Acest item are scopul de a identifica dacă instrumentul a fost proiectat cu intenția de a fi utilizat în condiții de administrare nesupravegheată sau necontrolată. Țineți cont de faptul că modurile de utilizare pot varia de la o versiune la alta a instrumentului. Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor, iar acuratețea sa trebuie verificată.</p> <p>Notă. Cele patru moduri sunt definite în <i>International Guidelines on Computer-Based and Internet Delivered Testing</i> (International Test Commission, 2005, pp. 5-6).</p>	<ul style="list-style-type: none"> <li>□ <i>Modul deschis</i>: Sesiunea de evaluare nu este supravegheată direct de către o persoană. De aceea, nu există o modalitate de autentificare a identității persoanei evaluate. Testele realizate pe internet, care nu solicită înregistrarea utilizatorilor, pot fi considerate un exemplu pentru acest mod de administrare.</li> <li>□ <i>Modul controlat</i>: Sesiunea de evaluare nu este supravegheată direct de către o persoană, însă testul este disponibil doar pentru utilizatori cunoscuți. Testele aplicate pe internet vor solicita utilizatorilor un username și o parolă. De regulă, aceste teste sunt proiectate să poată fi accesate doar o singură dată.</li> <li>□ <i>Modul supervizat (monitorizat)</i>: Condițiile în care este administrat testul sunt supravegheate într-o anumită măsură de o persoană. În acest mod, identitatea persoanelor evaluate poate fi autentificată. În cazul testelor aplicate pe internet, este nevoie ca un administrator să logheze un candidat și să confirme că testul a fost administrat și completat adecvat.</li> <li>□ <i>Modul dirijat (managed mode)</i>: Există un nivel ridicat de supraveghere și control asupra mediului în care se desfășoară testarea. În testarea computerizată (TC), acest nivel ridicat de control este de regulă atins prin faptul că se utilizează centre speciale de testare, unde accesul este controlat strict, securitatea este ridicată, există echipament și dotări de înaltă calitate, iar personalul care se ocupă de testare este calificat.</li> </ul>
2.11.	<p><b>Modul/modurile de administrare (selecțiați tot ce se aplică)</b></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor.</p>	<ul style="list-style-type: none"> <li>□ Administrare individuală interactivă</li> <li>□ Administrare supervizată de grup</li> <li>□ Aplicație computerizată instalată local – supervizată/monitorizată</li> </ul>

	<p>Orice alt echipament necesar, diferit de ce apare în lista de opțiuni (de ex., reportofon) trebuie descris aici. De asemenea, orice condiții speciale de testare trebuie descrise. Se presupune că există „condiții standard de testare” pentru o evaluare supravegheată/supervizată. Acestea ar include liniște, o cameră bine iluminată și bine ventilată, precum și un spațiu adecvat dotat cu birouri și scaune pentru evaluator/i și candidat/candidați.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Aplicație web pe calculator – supervizată/monitorizată</li> <li><input type="checkbox"/> Aplicație computerizată instalată local – nesupervizată/autoevaluare</li> <li><input type="checkbox"/> Aplicație web pe calculator – nesupervizată/autoevaluare</li> <li><input type="checkbox"/> Altceva (indicați):</li> </ul>
2.12.	<p><b>Timpul necesar pentru administrarea instrumentului</b> <i>(specificați pentru fiecare mod de administrare)</i></p> <p>Răspunsul la acest item trebuie formulat pe baza informațiilor furnizate de autor/editor.</p> <p>Răspunsul la acest element poate fi împărțit în mai multe componente. În majoritatea cazurilor, vor fi posibile mai degrabă estimări generale ale duratei, nu răspunsuri precise. Scopul este acela de a oferi potențialului utilizator o idee despre investiția de timp asociată cu utilizarea instrumentului. Nu includeți și timpul necesar pentru familiarizarea cu instrumentul în sine. Presupuneți că utilizatorul este experimentat și calificat.</p> <ul style="list-style-type: none"> <li>• Timpul de pregătire (timpul de care are nevoie persoana care administrează testul pentru a pregăti materialele necesare într-o sesiune de evaluare; timpul de acces și de logare pentru o administrare online).</li> <li>• Timpul de administrare pe sesiune: acesta include timpul necesar pentru completarea tuturor itemilor și o estimare a timpului necesar pentru instrucțiuni, parcurgerea unor itemi exemplu și discutarea oricăror</li> </ul>	<p>Pregătirea:</p> <p>Administrarea:</p> <p>Scorarea:</p>

	<p>comentarii la finalul sesiunii.</p> <ul style="list-style-type: none"> <li>• Scorarea: timpul necesar pentru a obține scorurile brute. În multe cazuri scorarea poate fi automată.</li> <li>• Analiza: timpul necesar pentru realizarea unor prelucrări asupra scorurilor brute, pentru a obține alte măsurători și pentru a produce o interpretare suficient de cuprinzătoare (presupunând că evaluatorul este familiar cu instrumentul). Din nou, această etapă poate fi automată.</li> <li>• Feedbackul: timpul necesar pentru a pregăti și a oferi feedback persoanei testate și altor părți interesate.</li> </ul> <p>Este cunoscut faptul că timpul necesar pentru ultimele două componente poate varia enorm, în funcție de contextul în care este folosit instrumentul. Totuși, unele indicații și comentarii vor fi de ajutor.</p>	<p>Analiza:</p> <p>Feedbackul:</p>
2.13.	<p><b>Indicați dacă sunt disponibile versiuni diferite ale instrumentului și ce versiune/versiuni este/sunt supusă/e acestei evaluări</b></p> <p>Precizați dacă există sau nu versiuni alternative ale instrumentului disponibil (forme autentice paralele sau pseudo-paralele, versiuni scurte, versiuni computerizate etc.) și descrieți aplicabilitatea fiecărei versiuni pentru diferite grupuri de persoane. În unele cazuri, diferite versiuni ale unui instrument sunt menite să fie echivalente una cu cealaltă – i.e. forme alternative. În alte cazuri, pot exista versiuni diferite pentru grupuri diferite (de ex., o versiune pentru copii și o versiune pentru adulți). În cazul în care există mai multe versiuni, indicați dacă acestea sunt echivalente/alternative sau dacă sunt concepute pentru a servi unor scopuri diferite - de ex. versiune scurtă</p>	

	<p>și lungă; versiune ipsativă și normativă. De asemenea, indicați dacă pot fi folosite sau nu părți ale testului în locul întregului instrument. Dacă există versiuni computerizate, descrieți pe scurt cerințele de software și hardware.</p> <p>Rețineți că dacă sunt disponibile teste computerizate (TC) și pachete online, ele trebuie menționate.</p>	
--	--	--

### 3. Măsurare și scorare

3.1.	<p><b>Procedura de scorare a testului</b> <i>(selectați tot ce se aplică)</i></p> <p>Acest item trebuie completat pe baza informațiilor oferite de autor/editor și a manualelor și documentației testului.</p> <p>Serviciile de consultanță (<i>bureau services</i>) sunt servicii oferite de furnizor - sau de un agent al furnizorului - pentru scorare și interpretare. În general, acestea sunt servicii opționale. Dacă scorarea și/sau interpretarea pot fi efectuate numai prin intermediul unui astfel de serviciu, atunci acest lucru ar trebui menționat, iar costurile aferente incluse la itemul privind costurile periodice.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Scorare computerizată, cu introducerea directă a răspunsurilor de către persoana evaluată</li> <li><input type="checkbox"/> Scorare computerizată, cu ajutorul unui Optical Mark Reader utilizat pentru formulare de răspuns în format hârtie</li> <li><input type="checkbox"/> Scorare computerizată, cu introducerea manuală a răspunsurilor din formularul de răspuns în format hârtie.</li> <li><input type="checkbox"/> Scorare manuală simplă - sunt necesare numai abilități administrative</li> <li><input type="checkbox"/> Scorare manuală complexă - necesită training prealabil</li> <li><input type="checkbox"/> Servicii de consultanță – de ex., scorare realizată de compania care vinde instrumentul</li> <li><input type="checkbox"/> Alt răspuns (vă rugăm să descrieți):</li> </ul>
3.2.	<p><b>Scorurile</b></p> <p>Acest item trebuie completat pe baza informațiilor oferite de autor/editor și a manualelor și documentației testului.</p> <p>Scurtă descriere a sistemului de scorare pentru obținerea scorurilor globale și parțiale, realizarea corecției pentru</p>	

	ghicitul răspunsurilor, generarea unei interpretări calitative etc.	
3.3.	<p><b>Scalele utilizate</b> (<i>selectați tot ce se aplică</i>)</p> <p>Acest item trebuie completat pe baza informațiilor oferite de autor/editor și a manualelor și documentației testului.</p>	<p><i>Scoruri bazate pe percentile</i></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Centile</li> <li><input type="checkbox"/> Clasificare în 5 trepte: împărțirea centilelor conform regulii 10:20:40:20:10</li> <li><input type="checkbox"/> Decile</li> <li><input type="checkbox"/> Alt răspuns (vă rugăm descrieți):</li> </ul> <p><i>Scoruri standard</i></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Scoruri Z</li> <li><input type="checkbox"/> Indicatori statistici pentru IQ, etc. (de ex., M= 100, SD=15 pentru Wechsler sau 16 pentru Stanford-Binet)</li> <li><input type="checkbox"/> Repere pentru testări educaționale standardizate (de ex., M=500, SD=100 pentru SAT)</li> <li><input type="checkbox"/> Stanine</li> <li><input type="checkbox"/> Stanine, scoruri C</li> <li><input type="checkbox"/> Scoruri T</li> <li><input type="checkbox"/> Alt răspuns (vă rugăm să descrieți):</li> </ul> <ul style="list-style-type: none"> <li><input type="checkbox"/> Scoruri critice, scoruri așteptate sau alți indicatori decizionali specifici</li> <li><input type="checkbox"/> Se utilizează doar scorul brut</li> <li><input type="checkbox"/> Alt răspuns (vă rugăm să descrieți):</li> </ul>
3.4.	<b>Transformarea scorurilor în scoruri standard</b>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Normalizate – scoruri standard obținute pe baza unor tabele de normalizare</li> <li><input type="checkbox"/> Ne-normalizate – scoruri standard obținute prin transformări liniare</li> <li><input type="checkbox"/> Nu se aplică</li> </ul>



#### 4. Rapoarte generate pe calculator

Această secțiune este pur *descriptivă*. Evaluarea rapoartelor trebuie realizată la secțiunea “Evaluare” a acestui formular.

Pentru situațiile în care sunt disponibile mai multe rapoarte generate, completați itemii 4.2 – 4.13 pentru fiecare raport sau secțiune individuală dintr-un raport. Acest sistem de clasificare poate fi utilizat pentru a descrie două rapoarte generate de un sistem - de exemplu, Raportul 1 poate fi destinat persoanelor evaluate sau altor utilizatori fără pregătire, în timp ce Raportul 2 poate fi destinat utilizatorilor cu pregătire, ce au competență în utilizarea instrumentului și înțeleg cum să interpreteze rezultatele.

4.1.	<b>Pentru acest instrument este posibilă generarea unor rapoarte computerizate?</b>  Dacă răspunsul la întrebare este "DA", atunci se aplică următoarea clasificare a tipurilor de rapoarte disponibile. În cazul multor instrumente este există o gamă variată de rapoarte. Vă rugăm să completați un formular separate pentru fiecare raport.	<input type="checkbox"/> Da (completați itemii de mai jos)  <input type="checkbox"/> Nu (mergeți la itemul 1 din secțiunea Condiții și costuri de livrare)
4.2.	<b>Denumirea sau descrierea raportului</b> <i>(vedeți introducerea la această secțiune)</i>	
4.3.	<b>Conținut</b> <i>(selectați tot ce se aplică)</i>  Rapoartele pot conține în întregime doar text sau text și reprezentări grafice sau tabele cu scoruri. În cazul în care există atât text, cât și date, cele două tipuri de conținut pot fi prezentate fie în paralel, fie grupate, în așa fel încât legătura dintre afirmațiile text și scoruri să fie clară.	<input type="checkbox"/> Doar text <input type="checkbox"/> Text și reprezentări grafice necorelate <input type="checkbox"/> Atât text cât și reprezentări grafice <input type="checkbox"/> Doar reprezentări grafice
4.4.	<b>Complexitate</b> <i>(selectați o variantă)</i>  Unele rapoarte sunt foarte simple, de exemplu presupun doar înlocuirea unui scor sten ( <i>standard tens</i> ) cu o unitate de text, pentru fiecare scală în parte. Altele sunt mai complexe și implică unități de text care corespund unor tipare sau	<input type="checkbox"/> Simplu (de ex., o listă cu descrieri ale scalelor) <input type="checkbox"/> Mediu (un amestec între descrieri simple și anumite descrieri bazate pe configurații)

	configurații la nivelul scorurilor și care iau în considerare efecte de interacțiune între scale.	<input type="checkbox"/> Complex (conține descrieri ale tiparelor și configurațiilor de la nivelul scorurilor și se ține cont de interacțiunile dintre scale)
4.5.	<p><b>Structura raportului</b> (<i>selectați o variantă</i>)</p> <p>Structura depinde de complexitate.</p>	<input type="checkbox"/> La nivel de scale – raportul este construit în jurul scalelor individuale. <input type="checkbox"/> La nivel factorial – raportul este construit în jurul factorilor de ordin înalt - precum 'Big Five' pentru evaluarea personalității. <input type="checkbox"/> La nivel de construct – raportul este construit în jurul unuia sau a mai multor seturi de legate de scorurile originale pe scale (de ex., în mediul organizațional, ar putea fi vorba despre tipuri de echipe, stiluri de leadership, sau toleranța la stres; în mediul clinic, acestea ar putea fi diferite tipuri de psihopatologie etc.) <input type="checkbox"/> Bazate pe criteriu – raportul este focalizat pe legătura cu rezultate empirice (de ex., performanța școlară, rezultatele terapeutice, performanța la locul de muncă, absenteism etc.) <input type="checkbox"/> Alt răspuns (vă rugăm să descrieți):
4.6.	<p><b>Sensibilitatea la context</b> (<i>selectați o variantă</i>)</p> <p>Oamenii elaborează rapoarte adaptând limbajul, forma și conținutul acestora în funcție de persoana care le va citi și țin cont de scopul evaluării și de contextul în care evaluarea are loc. Într-un context organizațional și de muncă, un raport generat pentru selecția de personal va fi diferit de un raport pentru orientare și dezvoltare; un raport pentru un manager de vârstă mijlocie va fi diferit de un raport pentru o persoană tânără aflată la începutul unui training, și așa mai departe. Într-un context educațional, un raport pentru evaluarea abilității generale de învățare și de funcționare într-un mediu de învățare a studenților va fi diferit de un raport pentru evaluarea măsurii în care un student are sau nu o</p>	<input type="checkbox"/> O singură versiune pentru toate contextele <input type="checkbox"/> Versiuni pre-definite în funcție de context; numărul contextelor: .... <input type="checkbox"/> Contexte definite de utilizator și rapoarte editabile

	<p>tulburare de învățare. Un raport direcționat către alți profesioniști cu scopul de a sugera obiective de învățare și modalități de intervenție va fi diferit de un raport dedicat părinților, prin care aceștia sunt informați în legătura cu punctele forte și cele slabe ale copiilor lor. Într-un context clinic, un raport produs în scop diagnostic va fi diferit de un raport în care se evaluează potențialul unui pacient de a adopta comportamente riscante. Un raport produs cu scopul de a oferi feedback pacienților va fi diferit de un raport produs cu scopul de a informa autoritățile cu privire la măsura în care este sigur ca un pacient să fie scos de la tratament involuntar.</p>	
4.7.	<p><b>Raportul clinic-actuarial</b> (<i>selectați tot ce se aplică</i>)</p> <p>Majoritatea sistemelor de raportare se bazează pe judecata clinică. Aceasta se referă la faptul că unul sau mai mulți oameni considerați „utilizatori experimentați” ai instrumentului vor scrie unitățile de text. Astfel, rapoartele vor încorpora interpretările lor specifice. Unele sisteme includ rapoarte actuariale – acestea se referă la faptul că afirmațiile sunt bazate pe studii empirice de validare ce corelează scorurile la scale cu variabile precum performanța la locul de muncă etc.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Bazat pe judecata clinică a unui expert</li> <li><input type="checkbox"/> Bazat pe judecata clinică a unui grup de experți</li> <li><input type="checkbox"/> Bazat pe relații empirice/statistice (actuariale)</li> </ul>
4.8.	<p><b>Posibilitatea de modificare</b> (<i>selectați o variantă</i>)</p> <p>Raportul generat este adesea fix. Însă unele sisteme pot genera un output sub forma unui fișier ce poate fi procesat de utilizator. Altele pot oferi acces la o variantă online interactivă atât pentru utilizator, cât și pentru persoana evaluată.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Nemodificabil (output fix ce permite doar printarea)</li> <li><input type="checkbox"/> Modificare limitată (limitată la anumite părți, de ex., pot fi editate doar câmpurile cu date biologice)</li> <li><input type="checkbox"/> Modificare nelimitată (de ex., accesând varianta Word a raportului)</li> <li><input type="checkbox"/> Raport interactiv ce oferă persoanei evaluate posibilitatea de a insera comentarii sau de a face aprecieri cu privire la acuratețea conținutului (de ex.,</li> </ul>

		prin oferirea accesului online la un motor interactive de raportare)
4.9.	<p><b>Stadiul de finalizare</b> (<i>selectați o variantă</i>)</p> <p>Măsura în care sistemul este conceput pentru a genera un text integrat - sub forma unui raport gata de a fi utilizat - sau o serie de "note", comentarii, ipoteze etc.</p>	<input type="checkbox"/> Pentru publicare <input type="checkbox"/> Schiță / draft
4.10.	<p><b>Transparență</b> (<i>selectați o variantă</i>)</p> <p>Sistemele diferă în ceea ce privește deschiderea sau transparența față de utilizator. Un sistem deschis este unul în care legătura dintre scorul la scală și text este neambiguă. O astfel de deschidere este posibilă doar dacă atât textul cât și scorurile sunt prezentate, și dacă legăturile dintre ele sunt explicite. Alte sisteme funcționează ca niște „cutii negre”, făcând dificil pentru utilizator să stabilească o relație între scoruri și text.</p>	<input type="checkbox"/> Conexiune clară între constructe, scoruri și text <input type="checkbox"/> Legătură ascunsă între constructe, scoruri și text <input type="checkbox"/> Amestec de legături clare/ascunse între constructe, scoruri și text
4.11.	<p><b>Stil și ton</b> (<i>selectați o variantă</i>)</p> <p>De asemenea sistemele diferă din punct de vedere al măsurii în care oferă ghidare și direcție persoanei care citește raportul. În context organizațional, o afirmație precum „Domnul X este foarte timid și nu ar fi un bun agent de vânzări...” este fermă, în timp ce alte afirmații sunt proiectate să sugereze ipoteze sau să ridice întrebări precum „Pe baza scorurilor înregistrate la scala Y, domnul X pare să fie o persoană foarte timidă comparativ cu un grup de referință format din agenți de vânzări. Dacă este corect, ar putea să i se pară dificil să lucreze într-un mediu axat pe vânzări. Acest aspect trebuie explorat mai în profunzime împreună cu dânsul.” Într-un</p>	<input type="checkbox"/> Directiv/ferm <input type="checkbox"/> Îndrumare/sugerează ipoteze <input type="checkbox"/> Alt răspuns (vă rugăm descrieți):

	<p>context educațional, o afirmație fermă ar putea fi: „Rezultatele arată că abilitățile matematice ale lui X sunt cu doi ani mai jos decât media colegilor săi”, în timp ce o afirmație proiectată să sugereze ipoteze ar putea fi: „Aceste rezultate indică faptul că X este devine distras ușor de stimuli externi în timp ce rezolvă o sarcină. Observațiile comportamentale din timpul testării susțin această posibilitate. Acest lucru ar trebui luat în considerare pentru proiectarea unui mediu de învățare potrivit pentru X”. Într-un context clinic, o afirmație fermă ar putea fi: „Scorurile la test indică faptul că pacientul manifestă neglijență vizuală severă și nu poate fi capabil să opereze în siguranță un vehicul cu motor”, în timp ce o afirmație proiectată să sugereze ipoteze ar putea fi: „Scorurile la test obținute de domnul X indică faptul că ar putea avea probleme în stabilirea relațiilor emoționale. Această posibilitate ar trebui explorată mai mult înainte de a formula o concluzie privind diagnosticul”.</p>	
4.12.	<p><b>Utilizatori</b> (<i>selectați tot ce se aplică</i>)</p> <p>În general, rapoartele sunt proiectate pentru a veni în întâmpinarea nevoilor uneia sau mai multor categorii de utilizatori. Utilizatorii pot fi împărțiți în patru grupuri principale:</p> <p><i>a) Utilizatori de test calificați.</i> Sunt persoane care dețin suficiente cunoștințe și abilități pentru a produce propriile rapoarte bazate pe scorurile la scale. Ar trebui să fie capabili să folosească rapoarte ce conțin terminologie tehnică din domeniul psihometriei și să realizeze conexiuni explicite între scale și descrieri. De asemenea, ar trebui să poată să personalizeze și să modifice rapoartele.</p>	<p>□ Utilizatori de test calificați</p>

	<p><i>b) Utilizatori de sistem calificați.</i> Deși nu pot să genereze propriile rapoarte pe baza unui set de scoruri, persoanele din această categorie dețin competența de a se folosi de outputurile emise de sistem. Nivelul trainingului necesar pentru atingerea acestei competențe va varia considerabil în funcție de natura rapoartelor computerizate (de ex., bazate pe trăsătură vs. bazate pe competență, simple sau complexe) și de modul în care aceste rapoarte vor fi utilizate (daca mizele sunt mici sau mari).</p> <p><i>c) Persoane evaluate.</i> Persoanele care completează testul, în general, nu au cunoștințe anterioare nici despre instrument, nici despre tipul raportului generat de sistem. Va fi nevoie ca rapoartele pentru astfel de persoane să fie scrise într-un limbaj ce nu necesită cunoștințe de psihometrie sau despre instrument.</p> <p><i>d) Terți.</i> Această categorie include persoane - altele decât candidatul – față de care informații prezentate în raport trebuie să rămână ascunse, sau care pot primi o copie a raportului. Pot fi potențiali angajatori, șeful persoanei evaluate, sau supervizorul ori părintele unui tânăr care primește consiliere în carieră. Limbajul necesar pentru persoanele din această categorie va fi similar cu cel necesar pentru rapoartele destinate persoanelor evaluate.</p>	<p><input type="checkbox"/> Utilizatori de sistem calificați</p> <p><input type="checkbox"/> Persoane evaluate</p> <p><input type="checkbox"/> Terți</p>
4.13.	<p><b>Distribuitorii oferă servicii de modificare a rapoartelor sau de elaborare a unor rapoarte computerizate personalizate?</b> (selecțai o variantă)</p>	<p><input type="checkbox"/> Da</p> <p><input type="checkbox"/> Nu</p>

## 5. Condiții și costuri de livrare

Această secțiune definește ce va livra autorul/editorul, cui, în ce condiții și cu ce costuri. Definește condițiile impuse de furnizor cu privire la cine poate sau nu poate să obțină materialele instrumentului. Dacă una dintre opțiuni nu este potrivită cu condițiile de livrare, oferiți o descriere a condițiilor relevante.

5.1.	<b>Documentația furnizată de distribuitor ca parte a pachetului testului</b> ( <i>selectați tot ce se aplică</i> )	<input type="checkbox"/> Manual de utilizare <input type="checkbox"/> Manual tehnic (psihometric) <input type="checkbox"/> Informații tehnice suplimentare și actualizări (de ex., etaloanele locale, studii locale de validare etc.) <input type="checkbox"/> Cărți și articole de interes
5.2.	<b>Forma publicării</b> ( <i>selectați tot ce se aplică</i> )  De exemplu, manualele tehnice pot fi actualizate și disponibile pentru descărcare de pe Internet, în timp ce manualele de utilizare sunt oferite în format hârtie sau pe CD/DVD.	<input type="checkbox"/> Hârtie <input type="checkbox"/> CD sau DVD <input type="checkbox"/> Download de pe internet <input type="checkbox"/> Altele (specificați):
<p>Itemii 5.3 – 5.5 se referă la costuri. Este foarte probabil ca aceste informații să își piardă cel mai rapid actualitatea. Se recomandă ca furnizorul sau autorul/editorul să fie contactați cât mai aproape posibil de momentul publicării evaluării, astfel încât răspunsurile la acești itemi să conțină informații cât mai actuale.</p>		
5.3.1	<b>Costuri de început</b>  Prețul unui set complet de materiale (toate manualele și alte materiale suficiente pentru cel puțin o administrare). Specificați numărul persoanelor care pot fi evaluate cu materialele achiziționate în contul costurilor de început și dacă aceste costuri includ și materiale pentru o evaluare repetată.  Acest item are menirea de a identifica costul inițial, de bază. Acesta se referă la costurile necesare pentru obținerea unui set complet de materiale de referință, chei de scorare și așa mai departe. Sunt incluse costurile de	

	<p>training dacă instrumentul este unul „închis” – adică necesită un cost de training de neevitat, indiferent de nivelul anterior de calificare al utilizatorului. În astfel de cazuri, includerea trainingului în costul de bază trebuie menționată explicit. Costurile inițiale NU includ costurile pentru echipament de uz general, precum calculatoare, DVD player ș.a. Totuși, necesitatea de a avea un astfel de echipament ar trebui menționată. În general, definiți: orice costuri speciale pentru training; costurile pentru manualul de utilizare; manualul/manualele tehnice; exemplarul sau setul de materiale de referință; costurile inițiale de software etc.</p>	
5.3.2	<p><b>Costuri periodice</b></p> <p>Specificați, dacă este cazul, costurile recurente de administrare și scorare separat de costurile de interpretare (a se vedea secțiunile 5.4.1 – 5.5).</p> <p>Acest item se referă la costul continuu de utilizare a instrumentului. Acesta include costul materialelor specifice instrumentului (foi de răspuns, foi de întrebări de unică folosință sau reutilizabile, foi pentru prezentarea profilului, coduri pentru aplicarea computerizată etc.) pe persoană pe administrare. În majoritatea cazurilor, pentru administrarea în format hârtie, astfel de materiale nu sunt disponibile într-un singur exemplar, ci tind să fie livrate în seturi de 10, 25 sau 50.</p> <p>Indicați taxele de licență anuale sau pe persoană (inclusiv codurile de utilizare a software-ului, dacă este cazul), costurile achizițiilor sau a materialelor reutilizabile pentru închiriere și costurile materialelor de unică</p>	



	folosință necesare per candidat.	
5.4.1	<b>Prețurile pentru rapoartele generate de softwareul instalat de utilizator</b>	
5.4.2	<b>Prețurile pentru rapoartele generate de serviciul prin poștă/fax</b>	
5.4.3	<b>Prețurile pentru rapoartele generate de serviciul prin Internet</b>	
5.5.	<b>Prețurile pentru alte servicii de birou: corectarea sau elaborarea rapoartelor automate</b>	
5.6.	<p><b>Calificările necesare pentru utilizarea testului, solicitate de furnizorul testului</b> (<i>selectați tot ce se aplică</i>)</p> <p>Acest item se referă la calificările pe care utilizatorul trebuie să le aibă, conform cerințelor furnizorului. Pentru acest item, în cazul în care autorul/editorul a oferit informații privind calificarea utilizatorului, acestea trebuie marcate conform categoriilor alăturate. În cazul în care cerințele de calificare nu sunt clare, acest lucru ar trebui menționat în rubrica „Altele”, nu la „Nicio calificare”. „Nicio calificare” înseamnă că există o afirmație explicită privind lipsa nevoi de calificare.</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Nicio calificare</li> <li><input type="checkbox"/> Acreditare specifică testului</li> <li><input type="checkbox"/> Acreditare pentru testarea generală a performanței: măsurarea performanței maxime de însușire a unor cunoștințe sau abilități (echivalent cu Nivelul 2 EFPA)</li> <li><input type="checkbox"/> Acreditare pentru testarea generală abilităților și aptitudinilor: măsurarea performanței maxime în relație cu potențialul de atingere (echivalent cu Nivelul 2 EFPA)</li> <li><input type="checkbox"/> Acreditare în evaluarea personalității și în evaluare generală: măsurarea comportamentelor tipice, a atitudinilor și a preferințelor (echivalent cu Nivelul 2 EFPA)</li> <li><input type="checkbox"/> Altele (specificați):</li> </ul>
5.7.	<p><b>Calificările profesionale necesare pentru utilizarea instrumentului</b> (<i>selectați tot ce se aplică</i>)</p> <p>Acest item se referă la calificările pe care utilizatorul trebuie să le aibă, conform cerințelor furnizorului. Pentru acest item, în cazul în care autorul/editorul a oferit informații privind calificarea utilizatorului, acestea trebuie marcate conform</p>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Nicio calificare</li> <li><input type="checkbox"/> Psiholog practician, specializat în domeniul de aplicabilitate specific testului</li> <li><input type="checkbox"/> Psiholog practician</li> <li><input type="checkbox"/> Psiholog cercetător</li> <li><input type="checkbox"/> Orice cercetător care lucrează în mediul academic</li> <li><input type="checkbox"/> Practician care activează în profesii asociate relevante (terapie, medicină, consiliere, educație, resurse umane etc.).</li> </ul> <p>Specificați: .....</p>

<p>categoriilor alăturate. În cazul în care cerințele de calificare nu sunt clare, acest lucru ar trebui menționat în rubrica „Altele”, nu la „Nicio calificare”. „Nicio calificare” înseamnă că există o afirmație explicită privind lipsa nevoi de calificare.</p> <p>Pentru detalii privind standardele EFPA de utilizare, consultați ultima versiune a acestora regăsite pe pagina web a EFPA.</p>	<ul style="list-style-type: none"> <li>□ Nivelul 1 EFPA pentru utilizatorii de teste sau echivalentul național</li> <li>□ Nivelul 2 EFPA pentru utilizatorii de teste sau echivalentul național</li> <li>□ Calificare de specialist, echivalentă cu standardele aferente Nivelului 3 EFPA pentru utilizatorii de teste.</li> <li>□ Altele (indicați):</li> </ul>
--	--

## PARTEA A DOUA – EVALUAREA INSTRUMENTULUI

### Sursele de informare

Există patru surse potențiale de informare, ce pot fi consultate pentru realizarea acestei evaluări:

**1. Manualul și/sau rapoartele oferite de autor/editor utilizatorului:**

Acestea sunt întotdeauna livrate de autor/editor/distribuitor înainte ca instrumentul să fie avizat de organizația care îl evaluează, și alcătuiesc materialele supuse evaluării.

**2. Informațiile accesibile în literatura de specialitate sau în alte surse:**

În general, aceste informații sunt identificate de către evaluator, iar acesta se poate folosi de ele în momentul evaluării. Instrumentul poate fi evaluat din punct de vedere al măsurii în care manualul include aceste informații.

**3. Informații deținute de autor/editor și care nu sunt publicate oficial sau distribuite:**

Distribuitorul/editorul/editura poate face ca aceste informații să fie disponibile încă de la început sau le poate trimite în momentul în care primește înapoi evaluarea, pentru verificarea acurateței anumitor afirmații sau fapte. Evaluatorul ar trebui să folosească aceste informații, dar să noteze foarte clar la începutul comentariilor referitoare la informațiile tehnice că „notele oferite în această evaluare se referă la materiale deținute de autor/editor/distribuitor, materiale ce [în mod normal] nu sunt furnizate utilizatorilor testului”. Dacă aceste materiale conțin informații valoroase, evaluarea generală ar trebui să conțină recomandarea ca autorul/editorul să publice aceste rapoarte și/sau să le pună la dispoziția cumpărătorilor testului.

**4. Informații confidențiale în scop comercial:**

În unele cazuri, editorii/editurile pot avea materiale importante din punct de vedere tehnic, dar pe care, din motive comerciale, nu doresc să le facă publice. În practică, sunt disponibile puține modalități de protecție a proprietății intelectuale a dezvoltatorilor de teste (ex., legea drepturilor de autor). Astfel de informații pot include rapoarte despre dezvoltarea unor algoritmi speciali de scorare, proceduri de generare a unor probe sau itemi, sau detalii despre tehnologia de generare a rapoartelor. În cazul în care conținutul acestor rapoarte este important pentru realizarea evaluării, asociația sau organizația responsabilă de evaluare ar putea să propună autorului/editorului întocmirea unui contract de confidențialitate. Acest contract ar fi încheiat între evaluatori și autor/editor. Apoi evaluatorul ar putea să evalueze și aceste informații confidențiale, să comenteze aspectele tehnice și să includă în evaluarea generală o precizare de tipul „notele marcate cu steluță se referă la materiale deținute de autor/editor sau de distribuitor, ce au fost examinate de evaluatori în baza unui acord de confidențialitate în scop comercial. Aceste informații nu sunt accesibile utilizatorilor”.

## Explicații privind notarea

Toate secțiunile sunt notate folosind sistemul de notare următor (consultați tabelul de pe pagina următoare). Sunt oferite descrieri detaliate pentru fiecare punctaj reper pentru evaluare.

Atunci când este oferită nota [0] sau [1] pentru o caracteristică a instrumentului considerată *esențială* pentru o utilizare sigură a acestuia, recomandarea va fi ca instrumentul să fie utilizat doar în circumstanțe excepționale, de către experți înalt calificați, sau în scop de cercetare.

În cadrul evaluării, trebuie indicat care sunt calitățile tehnice esențiale ale instrumentului, ținând cont de natura instrumentului și scopul pentru care a fost construit. Convenția recomandată este ca evaluările acestor calități esențiale să fie scrise cu caractere îngroșate (bold).

În secțiunile următoare, notele globale sau gradul de adecvare a informațiilor relevante pentru validitate, fidelitate și etaloane sunt marcate implicit cu caractere îngroșate (bold).

***Orice instrument notat cu una sau mai multe note de [0] sau [1] pentru caracteristici considerate esențiale pentru utilizarea sigură a instrumentului, ar trebui considerat sub standardele minime acceptate.***

Notă	Explicație*
[n/a]	Această caracteristică nu se aplică instrumentului de față
0	Evaluarea nu poate fi încadrată la "Nu" / absența caracteristicii, sau nu sunt oferite suficiente informații
1	Neadecvat
2	Adecvat
3	Bine
4	Excelent

\* EFPA definește o scală în cinci trepte, însă fiecare utilizator poate îmbina punctajele (de ex., să combine notele 3 și 4 într-o singură notă). Singura restricție este aceea de a stabili o distincție între neadecvat (sau prost) pe de o parte, și adecvat (sau bun) pe de altă parte. În cazul în care scala în cinci trepte este înlocuită sau personalizată, utilizatorul trebuie să stabilească o modalitate de legătură între punctaje/terminologie și scala în cinci trepte definită de EFPA.

## 6. Calitatea justificării utilității instrumentului, a modalității de prezentare și a informațiilor oferite

În această secțiune, este necesar să se acorde o serie de note pentru diferite aspecte sau caracteristici ale documentației care însoțește instrumentul (sau pachetul). Termenul „documentație” este folosit pentru a acoperi toate materialele furnizate sau disponibile ușor utilizatorilor calificați: de ex., manualul de administrare; manualele tehnice; broșurile cu etaloane; suplimentele la manual; actualizările realizate de autor/editor sau de furnizor etc.

Furnizorii sunt rugați să ofere câte un set complet cu aceste materiale pentru fiecare Evaluator.

### 6.1. Calitatea explicării utilității instrumentului

Dacă instrumentul este un test computerizat adaptiv (*computer-adaptive test*), trebuie acordată atenție deosebită itemilor 6.1.1 – 6.1.6.

Itemii ce trebuie cotați cu n/a sau cu 0 - 4		Notarea					
6.1.1	Fundamentarea teoretică a constructelor	n/a	0	1	2	3	4
6.1.2	Procedura de dezvoltare a testului (și/sau de traducere și adaptare)	n/a	0	1	2	3	4
6.1.3	Rigurozitatea analizei itemilor și modelul de analiză a itemilor	n/a	0	1	2	3	4
6.1.4	Prezentarea validității de conținut	n/a	0	1	2	3	4
6.1.5	Sinteza cercetărilor relevante	n/a	0	1	2	3	4
6.1.6	<b>Evaluarea globală a calității explicării utilității testului</b> Această evaluare globală se obține analizând notele oferite pentru itemii 6.1.1 – 6.1.5.	n/a	0	1	2	3	4

### 6.2. Adecvarea documentației disponibile utilizatorilor (manuale de utilizare și tehnice, etaloanele etc.)

Focusul acestei secțiuni este calitatea conținutului documentației accesibile utilizatorilor calificați. Subsecțiunea 2 se referă la complexitatea și claritatea documentației disponibile utilizatorilor (manualele de utilizare și cele tehnice, etaloanele etc.) ținând cont de informațiile acoperite și de explicații. Aspecte privind calitatea instrumentului reflectată de documentație sunt detaliate la: 6.1, 6.3, 9, 10 și 11.

Itemii ce trebuie cotați cu n/a sau 0-4; pentru calificativul „excellent” (4) sunt oferite repere.		Notarea					
6.2.1	Justificarea utilității (consultați notarea de la 6.1.6) Excelent: Există o descriere clară și logică a ceea ce instrumentul ar trebui să măsoare și a motivelor pentru care a fost proiectat în felul în care a fost proiectat.	n/a	0	1	2	3	4
6.2.2.1	Dezvoltarea Excelent: Sunt oferite detalii privind sursa itemilor, dezvoltarea materialului stimului în conformitate cu recomandările acceptate (de ex., Haladyna, Downing, & Rodriguez, 2002; Moreno, Martinez, & Muñiz, 2006), testarea pilot, analiza itemilor, studiile comparative și modificările făcute în timpul cercetărilor din etapa de dezvoltare.	n/a	0	1	2	3	4
6.2.2.2	Dezvoltarea testului prin traducere/adaptare Excelent: Informațiile din manual evidențiază faptul că procesul de traducere/adaptare a fost realizat în conformitate cu recomandările internaționale (ITC, 2000) și a inclus: <ul style="list-style-type: none"> <li>• Contribuția vorbitorilor nativi ai noii limbi</li> <li>• Revizuirii multiple făcute de persoane experte atât în limba, cât și în conținutul testului.</li> <li>• Traducere inversă din noua limbă în limba originală</li> <li>• Luarea în considerare a diferențelor culturale și lingvistice.</li> </ul>	n/a	0	1	2	3	4
6.2.3	Standardizarea Excelent: Sunt oferite informații clare și detaliate privind dimensiunea și proveniența eșantionului de standardizare și procedura de standardizare.	n/a	0	1	2	3	4
6.2.4	Etaloanele Excelent: Sunt oferite informații clare și detaliate privind dimensiunea și proveniența grupurilor normative, reprezentativitatea, condițiile de evaluare etc.	n/a	0	1	2	3	4

6.2.5	Fidelitatea Excelent: Sunt oferite explicații excelente privind fidelitatea și eroarea standard de măsurare (SEM). De asemenea, au fost testate consistența internă, fidelitatea test-retest și/sau fidelitatea formelor paralele și/sau fidelitatea inter-evaluator, iar erorile standard de măsurare obținute sunt însoțite de explicații privind relevanța lor și măsura în care instrumentul de evaluare este generalizabil.	n/a	0	1	2	3	4
6.2.6	Validitatea de construct Excelent: Validitatea de construct este explicată excelent, făcându-se referire la o gama largă de studii descrise clar și corect.	n/a	0	1	2	3	4
6.2.7	Validitatea de criteriu Excelent: Validitatea de criteriu este explicată excelent, făcându-se referire la o gama largă de studii descrise clar și corect.	n/a	0	1	2	3	4
6.2.8	Rapoartele generate de calculator Excelent: Sunt oferite informații clare și detaliate privind formatul, scopul, fidelitatea și validitatea rapoartelor generate de calculator.						
6.2.9	<b>Gradul de adecvare a documentației disponibile pentru utilizator (manuale de utilizare și tehnice, suplimentele cu etaloane etc.)</b> Această evaluare se obține analizând notele oferite pentru itemii 6.2.1 – 6.2.8.	n/a	0	1	2	3	4

### 6.3. Calitatea instrucțiunilor procedurale oferite utilizatorilor

Itemii ce trebuie cotați cu n/a sau 0-4; pentru calificativul „excelent” (4) sunt oferite repere.		Notarea					
6.3.1	<p>Pentru administrarea testului</p> <p>Excelent: Sunt oferite explicații și instrucțiuni procedurale clare și detaliate, împreună cu recomandări consistente și detaliate privind gestionarea întrebărilor candidaților și situațiile problemă.</p>	n/a	0	1	2	3	4
6.3.2	<p>Pentru scorarea testului</p> <p>Excelent: Sunt oferite informații clare și detaliate, împreună cu descrieri ale verificărilor necesare pentru gestionarea posibilelor erori de scorare.</p> <p>Dacă scorarea este făcută de calculator, se poate stabili dacă aceasta este corectă?</p>	n/a	0	1	2	3	4
6.3.3	<p>Pentru etalonare</p> <p>Excelent: Sunt oferite informații clare și detaliate, împreună cu descrieri ale verificărilor necesare pentru gestionarea posibilelor erori de etalonare.</p> <p>Dacă etalonarea este făcută de calculator, există dovezi cum că transformarea este corectă și că a fost folosit eșantionul normativ corect?</p>	n/a	0	1	2	3	4
6.3.4	<p>Pentru interpretare și raportare</p> <p>Excelent: Sunt oferite recomandări detaliate privind interpretarea unor scoruri diverse, înțelegerea măsurătorilor normative și gestionarea relațiilor dintre diferite scale, împreună cu exemple ilustrative și studii de caz ; de asemenea, sunt oferite recomandări privind gestionarea efectului pe care inconsistența răspunsurilor, stilurile de răspuns, falsitatea etc. ar putea să îl aibă.</p>	n/a	0	1	2	3	4



6.3.5	<p>Pentru feedback și debriefing dedicat persoanelor evaluate și altor persoane implicate</p> <p>Excelent: Sunt oferite recomandări detaliate privind modul de a oferi feedback candidaților; este inclusă utilizarea rapoartelor generate de calculator (dacă acestea sunt disponibile).</p>	n/a	0	1	2	3	4
6.3.6	<p>Pentru o bună practică din punct de vedere al corectitudinii și erorilor sociale</p> <p>Excelent: Sunt prezentare informații detaliate în legătură cu studiile privind erorile cauzate de diferențele etnice și de gen, împreună cu atenționări relevante privind utilizarea și generalizarea elementelor valide.</p>	n/a	0	1	2	3	4
6.3.7	<p>Restricții privind utilizarea</p> <p>Excelent: Sunt oferite descrieri clare despre cine ar trebui și cine nu ar trebui să fie evaluat, împreună cu justificări bine explicate pentru restricții (de exemplu, tipurile de dizabilități, nivelurile de alfabetizare necesare etc.)</p>	n/a	0	1	2	3	4
6.3.8	<p>Suport tehnic și software</p> <p>Excelent: În cazul testării computerizate (TC) sau a testării web (Tw), sunt oferite informații complete privind cerințele de browser, instalarea oricărui software necesar și utilizarea acestuia – de asemenea, sunt acoperite posibilele erori și diferențe de sistem -, iar disponibilitatea suportului tehnic este descrisă clar.</p>	n/a	0	1	2	3	4
6.3.9	<p>Referințe și materiale suport</p> <p>Excelent: Se fac referiri detaliate la publicații relevante din literatura de specialitate și sunt utilizate cross-referințe la alte materiale legate de instrumentul de evaluare.</p>	n/a	0	1	2	3	4
6.3.10	<p><b>Calitatea instrucțiunilor procedurale oferite utilizatorilor</b></p> <p>Această evaluare se obține analizând notele oferite pentru itemii 6.3.1 – 6.3.9.</p>	n/a	0	1	2	3	4

**Comentariile evaluatorilor pentru documentație:** (comentarii privind justificarea utilității instrumentului, calitatea prezentării și a informațiilor oferite)

## 7. Calitatea materialelor de testare

### 7.1. Calitatea materialelor aferente testelor creion-hârtie

(această secțiune poate fi sărită dacă nu se aplică)

Itemii ce trebuie cotați cu n/a sau 0-4		Notarea					
7.1.1	Calitatea generală a materialelor de testare (broșurile de testare, foi de răspuns, obiectele de testare, etc.)	n/a	0	1	2	3	4
7.1.2	Ușurința cu care persoana evaluată poate să înțeleagă sarcina de testare.	n/a	0	1	2	3	4
7.1.3	Claritatea și consistența instructajului persoanelor evaluate (inclusiv itemi exemplu și exerciții de antrenament).	n/a	0	1	2	3	4
7.1.4	Ușurința cu care persoanele evaluate pot oferi răspunsurile la test.	n/a	0	1	2	3	4
7.1.5	Calitatea formulării itemilor și claritatea conținutului grafic în cazul itemilor non-verbali.	n/a	0	1	2	3	4
7.1.6	<b>Calitatea materialelor aferente testelor creion-hârtie</b>	n/a	0	1	2	3	4

Această evaluare se obține analizând notele oferite pentru itemii 7.1.1 – 7.1.5.						
--	--	--	--	--	--	--

## 7.2. Calitatea materialelor aferente testării computerizate (TC) și a celei web (TW)

(această secțiune poate fi sărită dacă nu se aplică)

Itemii ce trebuie cotați cu n/a sau 0-4		Notarea					
7.2.1	Calitatea modului în care este proiectat softwareul (de ex., robustețea din punct de vedere al operării în momentul în sunt apăstate taste greșite, sau atunci când conexiunea la internet se întrerupe etc.).	n/a	0	1	2	3	4
7.2.2	Ușurința cu care persoana evaluată poate să înțeleagă sarcina de testare.	n/a	0	1	2	3	4
7.2.3	Claritatea și consistența instructajului oferit persoanelor evaluate (inclusiv itemi exemplu și exerciții de antrenament), și calitatea explicațiilor privind utilizarea softwareului și modalitatea de răspuns dacă testul este administrat de calculator.	n/a	0	1	2	3	4
7.2.4	Ușurința cu care persoanele evaluate pot oferi răspunsurile la test.	n/a	0	1	2	3	4
7.2.5	Calitatea interfaței pentru utilizatori.	n/a	0	1	2	3	4
7.2.6	Securitatea testului din punct de vedere al accesului neautorizat la itemi sau la răspunsuri.	n/a	0	1	2	3	4
7.2.7	Calitatea formulării itemilor și claritatea conținutului grafic în cazul itemilor non-verbali.	n/a	0	1	2	3	4
7.2.8	<b>Calitatea materialelor aferente testării computerizate (TC) și a celei web (TW)</b> Această evaluare se obține analizând notele oferite pentru itemii 7.2.1 – 7.2.7.	n/a	0	1	2	3	4

## **8. Etaloanele**

### **Instrucțiuni generale privind notarea acestei secțiuni**

Stabilirea criteriilor pentru evaluarea calității tehnice a unui instrument este greu de făcut. Acest paragraf oferă câteva repere privind modul în care pot fi atribuite valori calificativelor „neadecvat”, „adecvat”, „bun”, „excelent”. Însă natura instrumentului, domeniul în care este aplicat, calitatea datelor pe care se bazează etaloanele, precum și tipul deciziilor pentru care etaloanele vor fi folosite, ar trebui să influențeze modul în care sunt acordate notele. Pentru a da sens scorurilor brute obținute la test, se pot diferenția două modalități de scalare sau categorisire a lor (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). O primă modalitate este de a deriva o serie de scoruri scalate/ etaloane pe baza distribuției scorurilor brute ale unui eșantion de referință (interpretare bazată pe norme) (consultați secțiunea 8.1). A doua modalitate constă în derivarea unor standarde dintr-o gamă de competențe aparținând unui domeniu sau pe baza unor aspecte ce trebuie stăpânite (interpretare bazată pe domeniu) sau pot fi stabilite scoruri prag pe baza unor cercetări de validitate empirică (interpretare bazată pe criteriu) (consultați sub-secțiunea 8.2). Aplicând ultimele două variante, scorurile brute sunt categorisite în două sau mai multe intervale (de ex., repartizarea pacienților în diferite grupuri de tratament în funcție de intervalele în care se încadrează scorurile lor; aplicarea unui program educațional corectiv elevilor cu scoruri mai mici de o anumită valoare critică; acceptarea sau respingerea candidaților în procesul de selecție de personal).

## 8.1. Interpretarea bazată pe norme

*(Această subsecțiune poate fi sărită dacă nu se aplică)*

### Puncte esențiale privind normele internaționale

Trebuie să se acorde o atenție deosebită gradului de adecvare a normelor internaționale (aceeași limbă). În cazul în care acestea au fost atent stabilite pe baza unor eșantioane extrase de la nivelul mai multor țări, acestea ar trebui să fie evaluate la fel ca grupurile normative naționale (o singură limbă). În cazul în care este dată o normă non-locală, trebuie să fie oferite dovezi puternice atât pentru echivalența celor două versiuni ale testului, cât și pentru justificarea utilizării acestei norme. În general, astfel de dovezi presupun studii prin care să se demonstreze echivalența scalară între versiunea sursă și versiunea țintită. *Dacă astfel de detalii nu au fost raportate, atunci acest lucru ar trebui comentat în câmpul pentru comentariile evaluatorilor de la finalul secțiunii 8.*

O normă internațională ar putea fi cea mai potrivită pentru aplicabilitate internațională (i.e. compararea persoanelor care au dat testul în limbi diferite), însă pentru evaluarea acestei decizii, ar trebui luate în considerare aspectele listate mai jos. În general, pentru utilizarea unei norme internaționale, e necesară cel puțin demonstrarea echivalenței măsurării între versiunea sursă și versiunea țintită a testului.

#### ***Natura eșantionului***

- Omogenitatea eșantionului din punct de vedere al originii (de ex., un eșantion ce cuprinde 95% germani, 2% italieni și 3% englezi nu este un eșantion internațional autentic). Un eșantion poate fi calibrat în așa fel încât să reflecte cât mai bine caracteristicile relevante.
- Omogenitatea eșantionului din punct de vedere al istoricului participanților și al altor caracteristici situaționale (ocupație, educație, circumstanțele testării etc.). Eșantioanele normative care nu pot fi evaluate din acest punct de vedere nu sunt adecvate.

#### ***Tipul măsurării:***

- În cazul în care măsurarea presupune puțin conținut verbal sau absența unui conținut verbal, traducerea va fi afectată foarte puțin sau deloc. Acest lucru este valabil pentru testele de performanță și într-o anumită măsură pentru testele de raționament abstract și bazate pe diagrame, pentru care impactul asupra scorurilor ar trebui să fie mai mic.

#### ***Echivalența versiunilor testului aplicate pe eșantioane diferite din punct de vedere al limbii:***

- Trebuie să existe dovezi pentru faptul că toate versiunile sunt traduse/adaptate bine.
- Există dovezi pentru faptul că unul sau mai multe grupuri au completat testul într-o limbă diferită de limba natală?

#### ***Măsura în care scorurile obținute pe eșantioane diferite sunt asemănătoare:***

- Trebuie oferite date în legătură cu structura relativă a scorurilor înregistrate pe subeșantioane provenite din țări diferite. În cazul în care sunt diferențe mari, acestea trebuie luate în considerare, iar implicațiile asociate trebuie discutate. De ex., dacă pe un eșantion spaniol sunt obținute scoruri mai mari la o anumită scală comparativ cu un eșantion olandez, există vreo explicație despre ce înseamnă să compari membrii oricărui grup, sau un al treilea grup, cu media celor două? Există vreo interpretare pentru diferențe?

*Absența dovezilor/datelor de acest gen trebuie comentate la secțiunea Comentariile evaluatorilor de la finalul secțiunii.*

**Manualul instrumentului trebuie să conțină indicații cu privire la generalizarea etaloanelor dincolo de grupurile incluse pentru normele internaționale:**

- De ex., dacă eșantionul normativ conține 20% germani, 20% francezi, 20% italieni, 20% englezi și 20% olandezi, acesta poate fi folosit ca grup comparativ pentru candidați suedezi sau belgieni, dar nu poate fi folosit astfel pentru un grup de candidați chinezi.

8.1	<p><b>Interpretarea bazată pe norme</b></p> <p>În cazul în care un instrument este proiectat în așa fel încât utilizarea lui să nu presupună raportarea la etaloane sau la eșantioane normative (de ex., testele ipsative proiectate doar pentru comparații intra-individuale), trebuie folosită categoria „nu se aplică” (și nu cea „nu sunt oferite informații”). Totuși, evaluatorul trebuie să evalueze dacă absența unor etaloane este justificată; dacă nu, trebuie folosită categoria „nu sunt oferite informații”.</p>													
8.1.1	<p>Adecvarea pentru uz local, indiferent dacă există etaloane locale sau internaționale</p> <p>Țineți cont de faptul că, pentru testele adaptate, doar etaloanele locale (la nivel național) sau cele cu adevărat internaționale, sunt eligibile pentru notele 2, 3, sau 4, chiar dacă constructul s-a dovedit a fi echivalent croș-cultural. În cazul în care apar probleme din punct de vedere al invariantei măsurării, ar trebui oferite etaloane separate pentru (sub)grupuri, iar fiecare problemă trebuie explicată.</p> <table border="1" data-bbox="297 1045 1421 1623"> <tr> <td data-bbox="297 1045 1339 1119">Nu se aplică.</td> <td data-bbox="1339 1045 1421 1119">n/a</td> </tr> <tr> <td data-bbox="297 1119 1339 1192">Nu sunt oferite informații.</td> <td data-bbox="1339 1119 1421 1192">0</td> </tr> <tr> <td data-bbox="297 1192 1339 1266">Nu este relevant la nivel local (de ex., eșantioane străine nepotrivite).</td> <td data-bbox="1339 1192 1421 1266">1</td> </tr> <tr> <td data-bbox="297 1266 1339 1371">Eșantion/eșantioane locale care nu se potrivesc bine cu domeniul de aplicabilitate relevant, dar care pot fi folosite cu precauție.</td> <td data-bbox="1339 1266 1421 1371">2</td> </tr> <tr> <td data-bbox="297 1371 1339 1476">Eșantioane locale sau eșantioane internaționale relevante pentru aplicabilitatea dorită.</td> <td data-bbox="1339 1371 1421 1476">3</td> </tr> <tr> <td data-bbox="297 1476 1339 1623">Eșantioane locale la nivel de țară sau eșantioane internaționale relevante extrase din populații bine-definite, în conformitate cu domeniul de aplicabilitate relevant.</td> <td data-bbox="1339 1476 1421 1623">4</td> </tr> </table>		Nu se aplică.	n/a	Nu sunt oferite informații.	0	Nu este relevant la nivel local (de ex., eșantioane străine nepotrivite).	1	Eșantion/eșantioane locale care nu se potrivesc bine cu domeniul de aplicabilitate relevant, dar care pot fi folosite cu precauție.	2	Eșantioane locale sau eșantioane internaționale relevante pentru aplicabilitatea dorită.	3	Eșantioane locale la nivel de țară sau eșantioane internaționale relevante extrase din populații bine-definite, în conformitate cu domeniul de aplicabilitate relevant.	4
Nu se aplică.	n/a													
Nu sunt oferite informații.	0													
Nu este relevant la nivel local (de ex., eșantioane străine nepotrivite).	1													
Eșantion/eșantioane locale care nu se potrivesc bine cu domeniul de aplicabilitate relevant, dar care pot fi folosite cu precauție.	2													
Eșantioane locale sau eșantioane internaționale relevante pentru aplicabilitatea dorită.	3													
Eșantioane locale la nivel de țară sau eșantioane internaționale relevante extrase din populații bine-definite, în conformitate cu domeniul de aplicabilitate relevant.	4													
8.1.2	<p>Adecvarea pentru aplicațiile dorite</p> <table border="1" data-bbox="297 1686 1421 1831"> <tr> <td data-bbox="297 1686 1339 1770">Nu se aplică.</td> <td data-bbox="1339 1686 1421 1770">n/a</td> </tr> <tr> <td data-bbox="297 1770 1339 1831">Nu sunt oferite informații.</td> <td data-bbox="1339 1770 1421 1831">0</td> </tr> </table>		Nu se aplică.	n/a	Nu sunt oferite informații.	0								
Nu se aplică.	n/a													
Nu sunt oferite informații.	0													

	Etalonul sau etaloanele nu sunt adecvate pentru aplicațiile dorite.	1		
	Etaloane pentru populația generală și/sau tabele cu intervale normative adecvate, sau etaloane adecvate pentru o parte dintre aplicațiile dorite.	2		
	Tabele cu intervale normative bune.	3		
	O gamă excelentă de etaloane relevante, ce iau în considerare vârsta și genul; sunt incluse informații despre alte diferențe de la nivelul grupurilor (de ex., grup mixt din punct de vedere etnic).	4		
8.1.3	<p>Mărimea eșantioanelor (etalonarea clasică)</p> <p>Pentru majoritatea scopurilor, eșantioane mai mici de 200 de persoane testate sunt prea mici, având în vedere că intervalul de la extremele distribuției va fi mic. Eroarea standard a mediei pentru un scor <math>z</math> corespunzător unui eșantion de mărime <math>N = 200</math> este 0.071 din abaterea standard – sau doar mai bună decât un punct al unui scor <math>T</math>. Deși acest grad al lipsei de acuratețe poate avea doar consecințe minore la mijlocul distribuției, impactul asupra extremităților distribuției poate fi destul de mare (iar acestea ar putea fi scorurile cele mai relevante pentru deciziile luate pe baza rezultatelor la test). În general, dacă sunt etaloane internaționale, din cauza eterogenității, acestea trebuie să fie mai mari decât cerințele tipice pentru eșantioanele locale.</p> <p>Se folosesc diferite linii directoare atunci când mizele sunt mici și ridicate. În general, utilizarea ce implică mize mari se referă la situația în care o decizie însemnată se bazează cel puțin parțial pe scorul/scorurile obținute la test.</p>			
	Utilizarea cu miză mică	Decizii cu miză mare		
	Nu se aplică		n/a	
	Nu sunt oferite informații		0	
	Mărimea eșantionului inadecvată	de ex. < 200	de ex. 200-299	1
	Mărimea eșantionului adecvată	de ex. 200-299	de ex. 300-399	2
	Mărimea eșantionului bună	de ex. 300-999	de ex. 400-999	3
	Mărimea eșantionului excelentă	de ex. $\geq 1000$	de ex. $\geq 1000$	4
8.1.4	<p>Mărimea eșantioanelor pentru etalonare continuă</p> <p>Procedurile de etalonare continuă au devenit tot mai populare. Acestea sunt folosite în special pentru testele folosite în școli (de ex., clasele I-VIII) sau pentru anumite</p>			

	<p>intervale de vârstă (de ex., un test de inteligență pentru vârstele de 6-16 ani). Etalonarea continuă este mai eficientă deoarece, pentru a obține același nivel de acuratețe, sunt necesari mai puțin respondenți. Bechger, Hemker și Maris (2009) au calculat valori ale mărimii eșantioanelor utilizate în etalonarea continuă, în așa fel încât să fie asigurat același nivel de acuratețe ca în etalonarea clasică. Atunci când sunt folosite opt subgrupuri, <math>N = 70</math> (<math>8 \cdot 70</math>) conferă același nivel de acuratețe ca <math>N = 200</math> (<math>8 \cdot 200</math>). În aceste situații, acuratețea corespunzătoare etalonării continue este chiar mai bună pentru grupurile din centru, dar într-o anumită măsură mai redusă pentru grupurile din margine. Un alt avantaj, pe lângă eficiența mai crescută, este că valorile pentru grupurile normative intermediare pot fi calculate pe baza ecuației de regresie. Totuși, abordarea se bazează pe asumptii statistice mai degrabă stricte. Autorul testului trebuie să arate ca aceste asumptii au fost respectate, sau că abaterile de la aceste asumptii nu au consecințe serioase asupra acurateții etaloanelor.</p> <p>Rețineți că în cazul în care numărul eșantioanelor este mai mare, numărul respondenților din fiecare eșantion poate fi mai mic sau invers. Pentru deciziile cu impact mare, precum admiterea la școală, numărul necesar se mută cu o treaptă mai sus.</p>	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Mărimea eșantionului inadecvată (de ex., mai puțin de 8 subeșantioane, maxim 69 de respondenți în fiecare eșantion)	1
	Mărimea eșantionului adecvată (de ex., 8 subeșantioane, 70-99 de respondenți în fiecare eșantion)	2
	Mărimea eșantionului bună (de ex., 8 subeșantioane, 100-149 respondenți în fiecare eșantion)	3
	Mărimea eșantionului excelentă (de ex., 8 subeșantioane, cel puțin 150 de respondenți în fiecare eșantion)	4
8.1.5	<p>Proceduri pentru selectarea eșantionului (<i>selectați una</i>)</p> <p>Un eșantion normativ trebuie să fie reprezentativ pentru grupul la care se face referire. Un eșantion poate fi considerat reprezentativ pentru populația țintită atunci când componența sa este similară cu cea a populației respective, în funcție de un număr de variabile (de ex., vârstă, gen, educație), și atunci când eșantionul este format prin eșantionare probabilistică. Cu ajutorul acestei modalități de eșantionare, șansele de includere în eșantion sunt egale pentru fiecare membru al populației. Atât în eșantionarea probabilistică, cât și în cea nonprobabilistică, pot fi folosite mai multe metode.</p> <p>În eșantionarea probabilistică, atunci când un individ reprezintă o unitate de selecție, pot fi diferențiate trei metode: randomizare pură, selecție sistematică (de ex., fiecare al zecelea membru al unei populații) și stratificare (pentru unele variabile importante, precum genul, numerele ce urmează a fi selectate sunt stabilite în așa fel încât să fie</p>	



	<p>garantată reprezentativitatea acestor variabile). Totuși (de ex., din perspectiva eficienței), pot fi eșantionate și grupuri de persoane (de ex, clasele școlare), sau se poate aplica o combinație între eșantionarea la nivel de grup și cea la nivel de individ.</p> <p>În eșantionarea nonprobabilistică, pot fi diferențiate patru metode: eșantionare de conveniență (se adaugă în eșantion toate personale testate, așa cum se procedează pentru majoritatea eșantioanelor pentru selecția de personal; datele post-hoc pot fi clasificate în subeșantioane cu sens, pe baza informațiilor biografice și situaționale), eșantionarea cotă (se procedează ca în eșantionarea de conveniență, cu deosebirea că se specifică înainte numărul respondenților de care este nevoie în fiecare subeșantion, așa cum se întâmplă în cercetarea bazată pe chestionare), eșantionarea „bulgăre de zăpadă” (participanții sunt prieteni/cunoștințe, iar aceștia sunt rugați să găsească la rândul lor participanți în rândul prietenilor/cunoștințelor lor, etc.), și eșantionare cu scop (participanții fac parte din rândul grupurilor extreme).</p>	
	Nu sunt oferite informații	[ ]
	Eșantion probabilistic – randomizat	[ ]
	Eșantion probabilistic – sistematic	[ ]
	Eșantion probabilistic – stratificat	[ ]
	Eșantion probabilistic – clusterizat	[ ]
	Eșantion probabilistic – multifazic (de ex., primul cluster, urmat de randomizare la nivelul clusterelor)	[ ]
	Eșantion nonprobabilistic – de conveniență	[ ]
	Eșantion nonprobabilistic – cotă	[ ]
	Eșantion nonprobabilistic – „bulgăre de zăpadă”	[ ]
	Eșantion nonprobabilistic – cu scop	[ ]
	Altul, descrieți: .....	[ ]
8.1.6	Reprezentativitatea eșantionului/eșantioanelor normativ(e)	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Reprezentativitatea nu este potrivită pentru domeniul de aplicabilitate dorit sau reprezentativitatea nu poate fi stabilită adecvat pe baza informațiilor oferite	1
	Adecvată	2

	Bună	3
	Excelentă: Datele sunt colectate pe baza unei metode de eșantionare probabilistică; este oferită o descriere amănunțită a compoziției eșantionului/eșantioanelor și a populației/populațiilor, din punct de vedere al variabilelor biografice (de ex., gen, vârstă, educație, cultură, ocupație); pe baza acestor variabile se poate stabili o reprezentativitate bună	4
8.1.7	Calitatea informațiilor despre minoritate/diferențe asociate cu grupurile protejate, efectul vârstei, gen etc.	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Informații inadecvate	1
	Informații generale adecvate; există o analiză minimală	2
	Descrieri și analize bune ale grupurilor și diferențelor	3
	O gamă excelentă de analize și discuții privind aspectele relevante legate de utilizare și interpretare	4
8.1.8	Cât de vechi sunt studiile normative?	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Studii inadecvate, vechi de cel puțin 20 de ani	1
	Studii adecvate, etaloanele au o vechime de 15-19 ani	2
	Studii bune, etaloanele au o vechime de 10-14 ani	3
	Studii excelente, etaloanele au o vechime mai mică de 10 ani	4
8.1.9	Efectele de obișnuire/antrenament (sunt relevante doar pentru testele de performanță)	
	Nu se aplică	n/a
	Nu sunt oferite informații, deși este de așteptat să existe efecte de obișnuire/antrenament	[ ]
	Sunt oferite informații generale	[ ]

	Există etaloane pentru a doua aplicare a testului, după un interval test-retest obișnuit	[ ]
--	--	-----

## 8.2. Interpretare bazată pe criteriu

*(Această subsecțiune poate fi sărită dacă nu se aplică)*

Pentru determinarea scorurilor critice, se pot diferenția proceduri bazate pe judecata experților (aceste metode se mai numesc și etalonare cu referire la domeniu, consultați subcategoria 8.2.1) și proceduri ce presupun utilizarea datelor reale pentru a vedea relația dintre un rezultat la test și un criteriu extern (numite în sens restrâns cu referire la criteriu, consultați subcategoria 8.2.2).

8.2.1	Etalonare bazată pe domeniul de referință	
8.2.1.1	<p>Dacă este folosită judecata experților pentru a determina scorul critic, sunt aceștia selectați și pregătiți adecvat?</p> <p>Evaluatorii ar trebui să dețină cunoștințe privind domeniul de conținut al testului și ar trebui să aibă pregătire în evaluarea persoanelor care dau testul și a procedurii standard de stabilire a scorurilor prag. Procedura de selecție a evaluatorilor și de pregătire a lor trebuie descrisă.</p>	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Inadecvată	1
	Adecvată	2
	Bună	3
	Excelentă	4
8.2.1.2	<p>Dacă este folosită judecata experților pentru a determina scorul critic, numărul acestora este adecvat?</p> <p>Numărul necesar de evaluator depinde de sarcini și de context. Numerele sugerate ar trebui considerate un minim absolut.</p>	
	Nu se aplică	n/a
	Nu sunt oferite informații	0

	Inadecvat (mai puțin de doi evaluatori)	1
	Adecvat (doi evaluatori)	2
	Bun (trei evaluatori)	3
	Excelent (cel puțin patru evaluatori)	4
8.2.1.3	Dacă este folosită judecata experților pentru a determina scorul critic, ce procedură standard de stabilire a scorurilor prag este raportată? ( <i>selectați una</i> )	
	Nedelsky	[ ]
	Angoff	[ ]
	Ebel	[ ]
	Zieky și Livingston (grup limită)	[ ]
	Berk (grupuri de contrast)	[ ]
	Beuk	[ ]
	Hofstee	[ ]
	Alta, descrieți:	[ ]
8.2.1.4	Dacă este folosită judecata experților pentru a determina scorul critic, ce metodă de calculare a fidelității inter-evaluator este raportată?	
	Coeficientul $p_0$	[ ]
	Coeficientul Kappa	[ ]
	Coeficientul Livingston	[ ]
	Coeficientul Brennan și Kane	[ ]
	Coeficientul intra-clasă	[ ]
	Altul, descrieți:	[ ]
8.2.1.5	Dacă este folosită judecata experților pentru a determina scorul critic, care este mărimea coeficienților de fidelitate inter-evaluator (de ex., Kappa sau ICC)? În literatura de specialitate științifică nu există standarde clare pentru interpretarea acestor coeficienți, deși în general valorile sub .60 sunt considerate insuficiente. Mai	

	jos este urmărită clasificarea lui ShROUT (1998). Clasificarea trebuie utilizată cu o anumită precauție, pentru că prevalența sau rata de bază pot afecta valoarea lui Kappa.	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Inadecvată (de ex., $r < 0.60$ )	1
	Adecvată (de ex., $0.60 \leq r < 0.70$ )	2
	Bună (de ex., $0.70 \leq r < 0.80$ )	3
	Excelentă (de ex., $r \geq 0.80$ )	4
8.2.1.6	Cât de vechi sunt studiile normative?	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Studii inadecvate, vechi de cel puțin 20 de ani	1
	Studii adecvate, etaloanele au o vechime de 15-19 ani	2
	Studii bune, etaloanele au o vechime de 10-14 ani	3
	Studii excelente, etaloanele au o vechime mai mică de 10 ani	4
8.2.1.7	Efecte de obișnuire/antrenament (relevant doar pentru testele de performanță)	
	Nu sunt oferite informații, deși e de așteptat să existe efecte de obișnuire/antrenament	[ ]
	Sunt oferite informații generale	[ ]
	Există etaloane pentru a doua aplicare a testului, după un interval test-retest obișnuit	[ ]
<b>8.2.2</b>	<b>Etalonarea bazată pe criteriu</b>	
8.2.2.1	<p>Dacă scorul critic se bazează pe cercetări empirice, ce rezultate au fost obținute și care este calitatea acestor cercetări?</p> <p>Pentru a răspunde la această întrebare, nu pot fi oferite indicații explicite cu privire la ce înseamnă o relație acceptabilă. Aceasta nu doar pentru că ceea ce e considerat „mare” sau „mic” poate să fie diferit pentru fiecare criteriu prezis, dar și pentru că rezultatele predicției vor fi influențate de alte variabile, precum rata de bază sau</p>	

	prevalența. De aceea, evaluatorul trebuie să se bazeze pe expertiza sa. De asemenea, trebuie luate în considerare compoziția și mărimea eșantionului folosit pentru aceste cercetări (este acest eșantion similar cu grupul pe care îl țintește testul, este mai eterogen, sau mai omogen?).	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Inadecvate	1
	Adecvate	2
	Bune	3
	Excelente	4
8.2.2.2	Cât de vechi sunt studiile normative?	
	Nu se aplică	n/a
	Nu sunt oferite informații	0
	Studii inadecvate, vechi de cel puțin 20 de ani	1
	Studii adecvate, etaloanele au o vechime de 15-19 ani	2
	Studii bune, etaloanele au o vechime de 10-14 ani	3
	Studii excelente, etaloanele au o vechime mai mică de 10 ani	4
8.2.2.3	Efecte de obișnuire/antrenament (relevant doar pentru testele de performanță)	
	Nu sunt oferite informații, deși e de așteptat să existe efecte de obișnuire/antrenament	[ ]
	Sunt oferite informații generale	[ ]
	Există etaloane pentru a doua aplicare a testului, după un interval test-retest obișnuit	[ ]
8.3	<p><b>Adecvarea generală</b></p> <p>Această evaluare globală se obține analizând evaluările oferite la itemii 8.1 – 8.2.2.3. Evaluarea generală pentru <i>interpretarea bazată pe domeniul de referință</i> nu poate fi în niciun caz mai bună decât evaluarea la itemul privind mărimea eșantionului, dar poate fi mai slabă în funcție de celelalte informații. Dintre aceste informații, sunt relevante în special informațiile despre reprezentativitate și vechimea etaloanelor.</p>	

<p>Dacă sunt folosite eşantionare normative nonprobabilistice, calitatea etaloanelor poate să primească maxim calificativul „adecvat”, dar aceasta numai dacă descrierea eşantionului normativ reflecta faptul că distribuția variabilelor relevante este similară cu cea a grupului țintă sau la care se face referire. Evaluarea globală ar trebui să reflecte caracteristicile celor mai importante norme, mai degrabă decât media tuturor.</p> <p>Evaluarea globală pentru <i>interpretarea bazată pe criteriu</i>, pentru care au fost folosiți evaluatori pentru determinarea scorului critic, nu poate fi în niciun caz mai bună decât evaluarea nivelului de fidelitate inter-evaluator, dar poate fi mai slabă în funcție de celelalte informații oferite. Dintre aceste informații, sunt importante în special aplicarea corectă a metodei în cauză, precum și calitatea, pregătirea și numărul evaluatorilor. Dacă scorul critic este bazat pe cercetări empirice, evaluarea nu poate fi în niciun caz mai bună decât evaluarea pentru itemul 8.2.2.1, dar poate fi mai slabă, dacă studiile sunt prea vechi.</p>	
Nu se aplică	n/a
Nu sunt oferite informații	0
Inadecvat	1
Adecvat	2
Bine	3
Excelent	4

**Comentariile evaluatorilor cu privire la etaloane:** Scurt raport privind etaloanele și istoricul lor, inclusiv informații despre dispozițiile date de editură/editor sau autor pentru actualizarea etaloanelor în mod frecvent. Comentariile referitoare la normele non-locale ar trebui făcute aici.

## 9. Fidelitatea

### Instrucțiuni generale privind notarea acestei secțiuni

Fidelitatea se referă la măsura în care scorurile nu prezintă varianță a erorii de măsurare (ex., un interval de eroare de măsurare așteptată). Pentru fidelitate, instrucțiunile sunt bazate pe nevoia de a avea o eroare standard mică a estimărilor fidelității. Criteriile de instructaj pentru fidelitate sunt prezentate în legătură cu două contexte separate: utilizarea instrumentelor în luarea deciziilor despre grupuri de oameni (ex., diagnoză organizațională) și utilizarea lor pentru realizarea evaluărilor individuale. Cerințele fidelității sunt mai mari pentru cea de-a doua situație, decât pentru prima. Alți factori pot, de asemenea, afecta cerințele fidelității, precum tipul de decizie luată și dacă scalele sunt interpretate de sine stătător, sau agregate împreună cu alte scale într-o scală compozită. În cel de-al doilea caz, fidelitatea compozitului ar trebui să reprezinte focusul pentru notare, nu fidelitățile componentelor.

Atunci când un instrument a fost tradus și/sau adaptat pentru un context non-local, se pot aplica dovezi de fidelitate ale versiunii originale în sprijinul calității versiunii traduse/adaptate. În acest caz ar trebui propusă dovada echivalenței măsurii într-o nouă limbă față de cea originală. Pentru consistența internă, sunt preferabile dovezi ale fidelității bazate pe grupuri locale, deoarece aceste tipuri de dovezi sunt mai acurate și de obicei mai ușor de obținut. Pentru instrucțiuni cu privire la stabilirea echivalenței, vezi introducerea secțiunii despre Validitate. În Anexă este prezentat un ghid al punctelor critice în cazul unui instrument tradus și / sau adaptat dintr-un context non-local.

Aceste note oferă instrucțiuni legate de valorile care pot fi asociate cu o evaluare inadecvată, adecvată, bună sau excelentă, însă sunt intenționate doar ca ghid. Natura instrumentului, zona de aplicare, calitatea datelor pe baza cărora se estimează fidelitatea și tipurile de decizii pentru care urmează să fie folosit instrumentul ar trebui să influențeze felul în care se realizează evaluarea. În anumite condiții, o fidelitate de 0.70 este bună; în alte condiții, ar fi inadecvată. Din aceste considerente, judecata și expertiza reviewer-ului, sunt importante.

Pentru a oferi o idee legată de intervalul și distribuția valorilor asociate cu diferitele scale care compun un instrument, se introduce *numărul de scale* în fiecare secțiune. Spre exemplu, dacă un instrument utilizat pentru *decizii la nivel de grup* ar avea 15 scale, dintre care cinci au fidelitate retest mai scăzută de 0.6, șase între 0.60 și 0.70 și celelalte patru în intervalul 0.70-0.80, stabilitatea mediană poate fi considerată ca fiind "adecvată" (fiind categoria în care se situează / cade valoarea mediană de 15). Dacă este vizat mai mult de un singur studiu, trebuie calculată întâi valoarea mediană per scală, luând în considerare mărimile eșantionului; în anumite cazuri pot fi disponibile rezultate din meta-analize, care pot fi judecate în același fel. Acestea ar fi introduse ca:

Stabilitate	Număr de <i>scale</i> (dacă se aplică)	M*
Nicio informație oferită	[ - ]	0
Inadecvat (ex., $r < 0.60$ )	[ 5 ]	1
Adecvat (ex., $0.60 \leq r < 0.70$ )	[ 6 ]	2
Bun (ex., $0.70 \leq r < 0.80$ )	[ 4 ]	3
Excelent (ex., $r \geq 0.80$ )	[ 0 ]	4

\* M = stabilitate mediană



Pentru fiecare evaluare posibilă sunt oferite exemple de valori *exclusiv ca ghid* – mai ales distincțiile dintre ”Adecvat”, ”Bun” și ”Excelent”. Pentru decizii *cu miză mare*, precum în selecția de personal, aceste valori oferite ca exemple vor fi cu .10 mai mari. Cu toate acestea, merită menționat faptul că deciziile sunt deseori bazate pe scorurile scalelor agregate. Agregatele pot avea fidelități mult mai mari decât scalele lor componente primare. Spre exemplu, scalele primare din cadrul unui instrument multiscalar pot avea fidelități în jur de 0.70, în timp ce scale agregate secundare Big Five bazate pe acestea pot avea fidelități de 0.90. Manualele bune de test vor raporta atât fidelitățile scalelor secundare, cât și pe cele ale celor primare.

Deoarece cifrele mediane efective pot fi imposibil de calculat în multe cazuri, ceea ce se cere este cel mai bun estimat, conform informației oferite în documentare. Exista spațiu pentru a adăuga comentarii. Puteți nota aici orice preocupare aveți legat de acuratețea estimărilor dvs. Spre exemplu, în anumite cazuri, un nivel foarte înalt de consistență internă poate fi comentat ca indicând un ”specific umflat”.

<b>9</b>	<b>Fidelitate</b>	
9.1	Date oferite despre fidelitate ( <i>selectați două dacă se aplică</i> )	
	Nicio informație oferită	[ ]
	Doar un coeficient de fidelitate oferit (pentru fiecare scală sau subscală)	[ ]
	Doar un estimat al erorii standard de măsurare oferit (pentru fiecare scală sau subscală)	[ ]
	Coeficienți de fidelitate pentru un număr de grupuri diferite (pentru fiecare scală sau subscală)	[ ]
	Eroarea standard de măsurare oferită pentru un număr de grupuri diferite (pentru fiecare scală sau subscală)	[ ]
<b>9.2</b>	<b>Consistența internă</b>	
	<p>Utilizarea coeficienților de consistență internă nu este sensibilă în evaluarea fidelității testelor de viteză, scalelor heterogene (scalele empirice sau scalele criteriu-cheie; Cronbach, 1970), indicatorilor efectului (Nunnally &amp; Bernstein, 1994) și trăsăturilor emergente (Schneider &amp; Hough, 1995). În aceste cazuri, toți itemii referitori la consistența internă ar trebui marcați cu ”neaplicabil”. De asemenea, este eronată ca metodă de estimare a fidelității scalelor ipsative. Forme alternative sau măsuri de retest sunt mai potrivite pentru aceste tipuri de scale.</p> <p>Coeficienții de consistență internă oferă o estimare mai bună a fidelității decât coeficienții half-split corecțai cu formula Spearman-Brown. Așadar, utilizarea half-split este justificată doar atunci când, din diferite motive, informația despre răspunsuri la itemi individuali nu este disponibilă. Coeficienții half-split pot fi raportați în itemul 9.7 (Alte metode).</p>	
9.2.1	Mărimea eșantionului	

	Neaplicabil		n/a
	Nicio informație oferită		0
	Un studiu inadecvat (ex., eșantion mai mic de 100)		1
	Un studiu adecvat (ex., eșantion de 100-200)		2
	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime		3
	O gamă largă de studii / un număr mare de studii		4
9.2.2	Tipul de coeficienți raportați ( <i>selectați atâția câți sunt aplicabili</i> )		
	Neaplicabil		n/a
	Coeficient alpha sau KR-20		[ ]
	Lambda-2		[ ]
	Cea mai mare limită inferioară		[ ]
	Omega (analiză factorială)		[ ]
	Theta (analiză factorială)		[ ]
	Alții, descrieți: .....		[ ]
9.2.3	Mărimea coeficienților	Numărul de scale ( <i>dacă e aplicabil</i> )	M*
	Neaplicabil		n/a
	Nicio informație oferită	[ ]	0
	Inadecvat (ex., $r < 0.70$ )	[ ]	1
	Adecvat (ex., $0.70 \leq r < 0.80$ )	[ ]	2
	Bun (ex., $0.80 \leq r < 0.90$ )	[ ]	3
	Excelent (ex., $r \geq 0.90$ )	[ ]	4
9.2.4	Coeficienții de fidelitate sunt raportați cu eșantioane care .... ( <i>selectați una</i> )		
	.... Nu se potrivesc cu subiecții evaluați vizați, conducând la coeficienți mai favorabili (ex., creștere prin heterogenitate artificială)		[ ]

	.... Nu se potrivesc cu subiecții evaluați vizați, dar efectul asupra mărimii coeficienților este neclar	[ ]
	.... Nu se potrivesc cu subiecții evaluați vizați, conducând la coeficienți mai puțin favorabili (ex., scădere prin restricția intervalului)	[ ]
	.... Se potrivesc cu subiecții evaluați vizați	[ ]
	Inaplicabil	n/a
<b>9.3</b>	<b>Fidelitate test-retest – stabilitate temporală</b>	
	<p>Test-retest se referă la intervale relativ scurte de timp, în timp ce stabilitatea temporală se referă la intervale mai lungi, în care o mai mare schimbare este acceptabilă. Mai ales pentru testele utilizate pentru predicții pe perioade lungi de timp, ambele aspecte sunt relevante. Pentru a evalua stabilitatea temporală, poate fi necesar mai mult de un singur retest.</p> <p>Utilizarea unui design de tip test-retest nu este sensibilă pentru evaluarea fidelității măsurilor de stare (de fapt un coeficient mare de test-retest ar invalida caracterul de stare al testului). În acest caz, toți itemii referitori la fidelitatea test-retest ar trebui marcați ca "<i>Neaplicabili</i>".</p>	
9.3.1	Mărimea eșantionului	
	Neaplicabil	n/a
	Nicio informație oferită	0
	Un studiu inadecvat (ex., eșantion mai mic de 100)	1
	Un studiu adecvat (ex., eșantion de 100-200)	2
	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime	3
	O gamă largă de studii / un număr mare de studii	4
9.3.2	Mărimea coeficienților	Numărul de scale (dacă e aplicabil)
	Neaplicabil	n/a
	Nicio informație oferită	[ ]
	Inadecvat (ex., $r < 0.70$ )	[ ]
	Adecvat (ex., $0.70 \leq r < 0.80$ )	[ ]
		M*
		0
		1
		2

	Bun (ex., $0.80 \leq r < 0.90$ )	[ ]	3
	Excelent (ex., $r \geq 0.90$ )	[ ]	4
9.3.3	Date oferite despre intervalul de test-retest interval ( <i>selectați sau scrieți intervalul de test-retest</i> )		
	Neaplicabil		n/a
	Nicio informație oferită	[ ]	
	Intervalul este:		.....
9.3.4	Coeficienții de fidelitate sunt raportați cu eșantioane care .... ( <i>selectați una</i> )		
	.... Nu se potrivesc cu subiecții evaluați vizați, conducând la coeficienți mai favorabili (ex., creștere prin heterogenitate artificială)	[ ]	
	.... Nu se potrivesc cu subiecții evaluați vizați, dar efectul asupra mărimii coeficienților este neclar	[ ]	
	.... Nu se potrivesc cu subiecții evaluați vizați, conducând la coeficienți mai puțin favorabili (ex., scădere prin restricția intervalului)	[ ]	
	.... Se potrivesc cu subiecții evaluați vizați	[ ]	
	Neaplicabil		n/a

<b>9.4</b>	<b>Fidelitate echivalentă (paralelă sau forme alternative)</b>		
9.4.1	Mărimea eșantionului		
	Neaplicabil		n/a
	Nicio informație oferită		0
	Un studiu inadecvat (ex., eșantion mai mic de 100)		1
	Un studiu adecvat (ex., eșantion de 100-200)		2
	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime		3
	O gamă largă de studii / un număr mare de studii		4

9.4.2	Sunt îndeplinite asumptiile pentru paralelism* pentru versiuni diferite ale testului căruia i se investighează fidelitatea echivalentă? *Testele pot fi considerate teste/forme paralele dacă în cadrul aceluiași grup scorurile medii, varianțele și corelațiile cu alte teste sunt aceleași.		
	Neaplicabil		n/a
	Nicio informație oferită		0
	Inadecvat		1
	Adecvat		2
	Bun		3
	Excelent		4
9.4.3	Mărimea coeficienților	Numărul de scale (dacă e aplicabil)	M*
	Neaplicabil		n/a
	Nicio informație oferită	[ ]	0
	Inadecvat (ex., $r < 0.70$ )	[ ]	1
	Adecvat (ex., $0.70 \leq r < 0.80$ )	[ ]	2
	Bun (ex., $0.80 \leq r < 0.90$ )	[ ]	3
	Excelent (ex., $r \geq 0.90$ )	[ ]	4
9.4.4	Coeficienții de fidelitate sunt raportați cu eșantioane care .... (selectați una)		
	.... Nu se potrivesc cu subiecții evaluați vizați, conducând la coeficienți mai favorabili (ex., creștere prin heterogenitate artificială)		[ ]
	.... Nu se potrivesc cu subiecții evaluați vizați, dar efectul asupra mărimii coeficienților este neclar		[ ]
	.... Nu se potrivesc cu subiecții evaluați vizați, conducând la coeficienți mai puțin favorabili (ex., scădere prin restricția intervalului)		[ ]
	.... Se potrivesc cu subiecții evaluați vizați		[ ]
	Neaplicabil		n/a

<b>9.5</b>	<b>Metoda bazată pe TRI (Teoria răspunsului la item)<sup>3</sup></b>	
9.5.1	Mărimea eşantionului Este dificil de oferit un ghid uniform pentru adecvarea mărimii eşantioanelor în cazul metodelor TRI pentru estimarea fidelităţii, deoarece cerinţele sunt diferite în funcţie de formatul de răspuns la item şi de modelul de raspuns la item. În funcţie de modelul utilizat, valorile minime pentru mărimi adecvate ale eşantionului sunt: 200 pentru studii cu 1 parametru, 400 pentru studii cu 2 parametrii şi 700 pentru studii cu 3 parametrii (conform Parshall, Davey, Spray, & Kalohn, 2001). Aceste valori se aplică modelelor dihotomice, dar pot reprezenta un ghid pentru evaluator atunci când sunt utilizate modele politomice, pentru care mărimea eşantionului poate fi mai mică.	
	Neaplicabil	n/a
	Nicio informaţie oferită	0
	Un studiu inadecvat	1
	Un studiu adecvat	2
	Un studiu mare sau mai mult decât adecvat ca mărime	3
	O gamă largă de studii / un număr mare de studii	4
9.5.2	Tipuri de coeficienţi raportaţi ( <i>selectaţi pe cât de mulţi aplicabili</i> ) Prima metodă oferă fidelitatea trăsăturii latente estimate care, în TRI, înlocuieşte scorul adevărat estimat, de ex., scorul la test (vezi Embretson & Reise, 2000). A doua metodă este bazată pe informaţie despre itemi individuali şi oferă un estimat al fidelităţii atunci când cerinţele tipice pentru TRI sunt îndeplinite (Mokken, 1971). A treia metodă oferă un estimat al acurateţii măsurătorii în funcţie de poziţia trăsăturii latente.	
	Fidelitatea trăsăturii latente estimate	[ ]
	Rho	[ ]

<sup>3</sup> Metoda TRI explică şi analizează relaţia dintre caracteristicile individului şi răspunsurile la itemi, mai exact, relaţia dintre itemii testului şi probabilitatea alegerii răspunsului corect la acei itemi care solicită trăsăturile latente ale individului.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.

Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.

McDonald, R. P. (1999). *Test theory: A unified approach*. Mahwah, NJ: Lawrence Erlbaum.

Thissen, D., & Wainer, H. (Eds.). (2001). *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory*. New York, NY: Springer.

	Funcția informației	[ ]	
	Altele, descrieți:	[ ]	
	Neaplicabil	n/a	
9.5.3	<p>Mărimea coeficienților (bazată pe lungimea testului final)</p> <p>Atât instrucțiunile pentru coeficienții de fidelitate (inclusiv rho), cât și pentru funcția informației, sunt oferite. Instrucțiunile pentru funcția informației sunt bazate pe cele ale coeficienților de fidelitate din moment ce Informația = <math>1/SE^2</math>, și conform unor asumptii generale, <math>r = 1 - SE^2</math>. Luați aminte că SE și valorile informației sunt dependente de valoarea trăsăturii latente și că fiecare test are un interval în care valoarea informației este optimă. Evaluarea nu ar trebui să fie bazată în mod aprioric pe această valoare optimă, ci pe valoarea informației scorului sau a intervalului de scoruri care au importanță specifică (ex., scoruri critice). Pentru aceste scoruri, valoarea informației poate fi optimă, dar nu neapărat. Dacă nu există astfel de scoruri, evaluarea ar trebui să se bazeze pe valoarea medie a informației (vezi și Reise &amp; Havilund, 2005). Deoarece nu există multă experiență cu aceste procedee, recomandăm utilizarea lor cu precauție.</p>	<p>Număr de scale (dacă se aplică)</p>	M*
	Neaplicabil		n/a
	Nicio informație oferită	[ ]	0
	Inadecvat (ex., $r < 0.70$ ; informație $< 3.33$ )	[ ]	1
	Adecvat (ex., $0.70 \leq r < 0.80$ ; $3.33 \leq$ informație $< 5.00$ )	[ ]	2
	Bun (ex., $0.80 \leq r < 0.90$ ; $5.00 \leq$ informație $< 10.00$ )	[ ]	3
	Excelent (ex., $r \geq 0.90$ ; informație $\geq 10.00$ )	[ ]	4

<b>9.6</b>	<b>Fidelitate inter-evaluator</b>		
	<p>Dacă scorarea unui test nu implică procese de judecată (ex., simpla sumă a scorurilor unor itemi cu răspuns multiplu), acest tip de fidelitate ar trebui marcat ca "Neaplicabil". Deși fidelitatea inter-evaluator s-ar putea să nu se aplice unui test per total, se poate aplica unuia sau mai multor subteste (ex., anumite subteste ale unui test de inteligență).</p>		
9.6.1	Mărimea eșantionului		
	Neaplicabil		n/a
	Nicio informație oferită		0
	Un studiu inadecvat (ex., eșantion mai mic de 100)		1
	Un studiu adecvat (ex., eșantion de 100-200)		2
	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime		3
	O gamă largă de studii / un număr mare de studii		4
9.6.2	Tipul de coeficienți raportați ( <i>selectați atâția câți sunt aplicabili</i> )		
	Neaplicabil		n/a
	Procentaj de agreere		[ ]
	Coeficient Kappa		[ ]
	Corelație Intra-Clasă		[ ]
	Coeficient Iota		[ ]
	Altul, descrie:		[ ]
9.6.3	Mărimea coeficienților	Număr de scale (dacă se aplică)	M*
	Pentru anumite metode menționate la 9.6.2 alocarea punctajelor s-ar putea să nu fie aplicabilă, deoarece nu există niciun $r$ calculat.		
	Neaplicabil		n/a
	Nicio informație oferită	[ ]	0
Inadecvat (ex., $r < 0.60$ )	[ ]	1	



	Adecvat (ex., $0.60 \leq r < 0.70$ )	[ ]	2
	Bun (ex., $0.70 \leq r < 0.80$ )	[ ]	3
	Excelent (ex., $r \geq 0.80$ )	[ ]	4
<b>9.7</b>	<b>Alte metode de estimare a fidelității</b>		
9.7.1	Mărimea eșantionului		
	Neaplicabil		n/a
	Nicio informație oferită		0
	Un studiu inadecvat (ex., eșantion mai mic de 100)		1
	Un studiu adecvat (ex., eșantion de 100-200)		2
	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime		3
	O gamă largă de studii / un număr mare de studii		4
9.7.2	Descrieți metoda:		
9.7.3	Rezultate	Număr de scale (dacă se aplică)	M*
	Neaplicabil		n/a
	Nicio informație oferită	[ ]	0
	Inadecvat	[ ]	1
	Adecvat	[ ]	2
	Bun	[ ]	3
	Excelent	[ ]	4

9.8	<p><b>Adecvare generală</b></p> <p>Această evaluare generală se obține utilizând evaluările oferite itemilor 9.1 – 9.7.3. <i>Nu faceți o simplă medie a numerelor pentru a obține evaluarea generală.</i></p> <p>Pentru anumite instrumente, consistența internă poate fi inadecvată (trăsături largi sau agregate de scale), caz în care trebuie pus mai mult accentul pe datele de retest. În alte cazuri (măsuri ale stării), fidelitățile retest ar fi inadecvate, așadar accentul ar trebui pus pe consistențele interne. Pentru luarea deciziei finale, ar trebui să luați în considerare:</p> <ul style="list-style-type: none"> <li>– Dacă testul este utilizat în evaluare individuală sau pentru a lua decizii pentru grupuri de oameni</li> <li>– Natura deciziei (miză înaltă vs. miză scăzută)</li> <li>– Dacă unul sau mai multe (tipuri de) studii de fidelitate sunt raportate</li> <li>– Dacă sunt oferite și erorile standard de măsurare</li> <li>– Probleme procedurale, ex., mărimea grupului, număr de studii de fidelitate, eterogenitatea grupului/grupurilor pentru care sunt computați coeficienții, numărul de evaluatori dacă este computată decizia inter-evaluator, lungimea intervalului test-retest, etc.</li> <li>– Gradul de comprehensivitate al raportării în studiile de fidelitate.</li> </ul>										
	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">Nicio informație oferită</td> <td style="text-align: center; padding: 5px;">0</td> </tr> <tr> <td style="padding: 5px;">Inadecvat</td> <td style="text-align: center; padding: 5px;">1</td> </tr> <tr> <td style="padding: 5px;">Adecvat</td> <td style="text-align: center; padding: 5px;">2</td> </tr> <tr> <td style="padding: 5px;">Bun</td> <td style="text-align: center; padding: 5px;">3</td> </tr> <tr> <td style="padding: 5px;">Excelent</td> <td style="text-align: center; padding: 5px;">4</td> </tr> </table>	Nicio informație oferită	0	Inadecvat	1	Adecvat	2	Bun	3	Excelent	4
Nicio informație oferită	0										
Inadecvat	1										
Adecvat	2										
Bun	3										
Excelent	4										

**Comentariu al evaluatorilor despre fidelitate:** Subliniați aspectele puternice și slabe ale dovezilor disponibile despre fidelitate. Comentarii cu privire la generalizarea echivalenței/fidelității ar trebui făcute tot aici (dacă se aplică).

## 10. Validitatea

### Ghid general de evaluare a acestei secțiuni

Validitatea este măsura în care un test își îndeplinește scopul: se pot trage concluziile stabilite inițial din scorurile testării? În literatură, sunt diferențiate mai multe tipuri de validitate, de ex., Drenth și Sijtsma (2006, p. 334 - 340) menționează opt tipuri diferite. Diferențele pot avea legătură cu scopul validării sau cu procesul de validare prin tehnici specifice de analiză a datelor. În ultimele decenii ale secolului trecut a existat un consens în creștere, potrivit căruia validitatea ar trebui considerată ca un concept unitar și că diferențierile în tipurile de validitate ar trebui considerate ca modalități diferite doar de a strânge dovezi (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Borsboom, Mellenbergh și Van Heerden (2004) afirmă că un test este valabil pentru măsurarea unui atribut dacă variația atributului provoacă cauzal variații în rezultatele măsurate. Deși aceasta este o abordare diferită, și în opinia acestor autori o diferențiere între tipuri de validitate nu este relevantă.

Cu toate acestea, oricare ar fi abordarea preferată a validității, pentru o judecată standardizată, este necesară o minimă structurare a conceptului de validitate. Din acest motiv, subsecțiunile separate privind validitatea constructului și a criteriului sunt diferențiate. În funcție de scopul testului, unul dintre aceste aspecte ale validității poate fi mai relevant decât celălalt. Cu toate acestea, validitatea constructului este conceptul fundamental, iar dovezile privind validitatea criteriului pot contribui la stabilirea validității de construct a unui test.

Se constată, de asemenea, că un test poate avea validități diferite, în funcție de tipul deciziilor luate cu ajutorul testului, de tipul probelor utilizate, etc. Cu toate acestea, în cadrul unui proces de evaluare a testului, o judecată de calitate se realizează cu privire la validitatea (de construct sau de criteriu) a unui test. Această judecată ar trebui să reflecte calitatea dovezilor care susțin afirmația că testul poate fi folosit pentru interpretările menționate în manual. Cu cât aplicațiile intenționate sunt mai ample, cu atât mai multe dovadă de validitate ar trebui să ofere autorul/editorul. Rețineți că evaluarea finală pentru validitatea de construct și de criteriu va fi o medie a acestor dovezi și că pot exista situații sau grupuri pentru care testul poate avea valori mai mari sau mai mici (sau pentru care validitatea poate să nu fi fost studiată deloc).

Atunci când un instrument a fost tradus și / sau adaptat dintr-un context non-local, ar trebui propusă o dovadă a echivalenței măsurii într-o limbă nouă a originalului. Fără aceasta, nu este posibilă generalizarea rezultatelor dintr-o țară / limbă în alta. Exemple de dovezi echivalente:

- Invarianța în structura de construct – ex., prin structura factorială sau prin corelarea cu măsurile standard.
- Validitate similară a criteriului – ex., un profil similar al corelațiilor unui instrument multiscalar cu criteriile externe independente - cum ar fi evaluările competențelor profesionale.
- Itemii prezintă modele similare de încărcare a scalelor, ex., elementele se corelează în același model cu alte scale; itemii cei mai puternic / slab încărcăți sunt similari în limbile originale și noi.
- Candidații bilingvi au profiluri similare în două limbi (cf. fidelității formei alternative).

Generalizarea validității necesită dovezi mai puternice în traducerea testelor în cadrul familiilor lingvistice (ex., de la o limbă indo-europeană, la o limbă semită). Într-o astfel de situație,

echivalența este amenințată mai mult din cauza diferențelor dintre structura lingvistică și diferențele culturale. Cu toate acestea, generalizarea validității poate fi dedusă din dovezi ale invarianței validității în traduceri anterioare, atunci când un test a fost tradus în mai multe limbi. De exemplu, dacă un test suedez a fost deja tradus în franceză, germană și italiană și s-a dovedit a fi echivalent în aceste limbi.

În considerarea echivalenței, poate fi util de urmat clasificarea lui Van de Vijver și Poortinga (2005):

- Echivalență structurală / funcțională
  - Există dovezi că limbile sursă și țintă măsoară aceleași constructe psihologice în cadrul grupurilor. Acest lucru este demonstrat, în general, prin faptul că modelele de corelații între variabile sunt identice între grupuri.
- Echivalența unității de măsură
  - Există dovezi că unitățile de măsură sunt aceleași, dar există origini diferite între grupuri (diferențele individuale găsite în grupa A pot fi comparate cu diferențele găsite în grupa B, dar scorurile brute absolute pentru A și B nu sunt direct comparabile fără nici o formă de redimensionare).
- Echivalența Scalară / Scorului complet
  - Aceeași unitate de măsură și aceeași origine (de ex., scorurile brute au aceeași semnificații și pot fi comparate între grupuri).

Indicii de referință și notele din subsecțiunile 10.1 și 10.2 oferă unele îndrumări cu privire la valorile care trebuie asociate cu evaluări necorespunzătoare, adecvate, bune și excelente. Cu toate acestea, acestea sunt destinate să acționeze doar ca ghiduri. Natura instrumentului, domeniul său de aplicare, calitatea datelor pe care se bazează estimările validității și tipurile de decizii pentru care vor fi utilizate ar trebui să afecteze modul în care sunt realizate evaluările. Pentru validitate, liniile directe privind dimensiunile eşantioanelor se bazează pe analiza puterii pentru dimensiunile eşantioanelor necesare pentru a găsi validități de dimensiuni moderate, dacă acestea există.

### **10.1. Validitatea de construct**

Scopul validării constructului este de a găsi un răspuns la întrebarea dacă testul măsoară efectiv constructul dorit, parțial sau total, etc. Metodele comune pentru investigarea validității constructului sunt analiza factorială exploratorie sau confirmatorie, corelațiile item-test, compararea scorurilor medii ale grupurilor pentru care se pot aștepta diferențe de scor, testarea pentru invarianța structurii factorilor și item-bias (DIF) pentru diferite grupuri, corelații cu alte instrumente care sunt destinate măsurării aceluiași aspect (validitate convergentă) sau a diferitelor constructe (validitate discriminatorie), cercetări Multi-Trăsătură-Multi-metodă (MTMM), metodologii TRI și design-uri (cvasi-) experimentale.

<b>10.1</b>	<b>Validitatea de construct</b>	
10.1.1	Design-uri utilizate ( <i>selectați atâtea câte se aplică</i> )	
	Nicio informație oferită	[ ]
	Analiză Factorială Exploratorie	[ ]
	Analiză Factorială Confirmatorie	[ ]
	Corelații (Corectate) item-test	[ ]
	Testarea pentru invarianța structurii și funcționare diferențiată a itemilor pentru diferite grupuri	[ ]
	Diferențe între grupuri	[ ]
	Corelații cu alte instrumente și criterii de performanță	[ ]
	Corelații MTMM	[ ]
	Metodologie TRI	[ ]
	Design-uri (Cvasi-)Experimentale	[ ]
	Altul, descrie:	[ ]
10.1.2	Suportă rezultatele analizei factoriale (confirmatorie sau exploratorie) structura testului?	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4
10.1.3	Itemii se corelează suficient de bine cu scorul (sub) testului? Corelațiile foarte mari pot însemna faptul că elementele sunt mai mult sau mai puțin sinonime, iar conceptul măsurat poate fi foarte restrictiv (așa-numitul "specific umflat")	
	Nicio informație oferită	0

	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4
10.1.4	<p>Structura factorială este invariabilă în grupuri și/sau nu există item-bias (DIF)?</p> <p>Acest tip de cercetare poate fi realizat pe baza modelelor din cadrul teoriei clasice de testare sau al cadrului TRI. Dacă se constată item-bias, este necesar să se aprecieze efectul asupra scorului total (se acceptă efecte mici).</p>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4
10.1.5	<p>Există diferențe în scorurile medii între grupurile relevante, cum era de așteptat?</p> <p>De ex., elevii din clasa a VIII - a se așteaptă să aibă scoruri mai mari decât elevii din clasa a VI - a la un test de competență numerică; copiii cu diagnosticul de ADHD ar trebui să obțină un scor mai mare la un test de hiperactivitate, decât copiii care nu sunt diagnosticați cu ADHD; agenții de vânzări ar trebui să aibă scoruri mai mari la un test pentru cunoștințele comerciale, decât media populației lucrătoare. Chiar dacă rezultatele sunt în direcția așteptată, acest tip de cercetare este de obicei neconcludent în ceea ce privește validitatea de construct a testului. Cu toate acestea, valoarea acestui tip de cercetare este că, atunci când diferențele așteptate nu sunt prezentate, acest lucru ar ridica îndoieli puternice cu privire la validitatea de construct a testului.</p>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4

10.1.6	<p>Mediana și intervalul de corelații între test și teste care măsoară constructe similare</p> <p>Un element esențial al procesului de validare a constructului este corelarea scorului (scorurilor) testului cu scale de la instrumente similare, așa-numita validitate congruentă sau convergentă. Orientările privind coeficienții de validitate congruentă trebuie interpretați în mod flexibil. În cazul în care au fost corelate două instrumente foarte asemănătoare (cu datele obținute simultan), ne-am aștepta să găsim corelații de 0,60 sau mai mult pentru "adecvate". În cazul în care instrumentele sunt mai puțin similare, sau sesiunile de administrare sunt separate de un anumit interval de timp, valorile mai mici pot fi adecvate. Atunci când se evaluează validitatea congruentă, trebuie acordată atenție la interpretarea corelațiilor foarte mari. Atunci când corelațiile sunt peste 0,90, probabilitatea este ca scalele în cauză să măsoare exact același construct. Aceasta nu este o problemă dacă scalele în cauză reprezintă o nouă scală și un marker stabilit. Ar fi totuși o problemă, dacă scala (ele) în cauză ar fi fost proiectate astfel încât să adauge o varianță utilă la ce măsoară deja alte scale. Instrucțiunile oferite privesc corelațiile care nu sunt ajustate pentru varianța sau atenuarea metodei comune. Prin urmare, la luarea în considerare a coeficienților de congruență de validitate, trebuie luată în considerare și fidelitatea ambelor instrumente. De exemplu, atunci când ambele instrumente au o fidelitate de .75, corelația maximă dintre instrumente este .56. Dacă fidelitățile sunt mai mari, se așteaptă o corelație mai mare.</p>	
	Nicio informație oferită	0
	Inadecvat ( $r < 0.55$ )	1
	Adecvat ( $0.55 \leq r < 0.65$ )	2
	Bun ( $0.65 \leq r < 0.75$ )	3
	Excelent ( $r \geq 0.75$ )	4
10.1.7	Corelațiile cu alte instrumente demonstrează validitate discriminatorie bună în ceea ce privește constructele pe care testul nu ar trebui să le măsoare?	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4



10.1.8	<p>Dacă se utilizează un design Multi-Trăsătură-Multi-Metodă, rezultatele ajută la construirea validității testului (măsoară cu adevărat ce ar trebui să măsoare și nu altceva)?</p> <p>Rețineți că, dacă se utilizează un design MTMM, este posibil ca cercetarea menționată la punctele 10.1.6 și 10.1.7 să nu mai fie necesară.</p>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4
10.1.9	<p>Altul, de ex., metodologie TRI, design (cvasi-) experimental (descrieți):</p>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4
10.1.10	<p>Mărimea eșantionului</p> <p>Instrucțiunile de mai jos se referă la studii ce se încadrează în teoria clasică a testului. Pentru estimarea itemilor-parametru în cadrul metodologiei TRI, dimensiunile eșantioanelor "adecvate" sunt: mai mult de 200 pentru studii cu 1 parametru, mai mult de 400 pentru studii cu 2 parametri și mai mult de 700 pentru studii cu trei parametri (bazate pe Parshall, Davey, Spray, și Kalohn, 2001).</p>	
	Neaplicabil	0
	Nicio informație oferită	1
	Un studiu inadecvat (ex., eșantion mai mic de 100)	2
	Un studiu adecvat (ex., eșantion de 100-200)	3

	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime	4
10.1.11	<b>Calitatea instrumentelor ca și criterii sau markeri</b>	
	Nicio informație oferită	0
	Calitate inadecvată	1
	Calitate adecvată	2
	Calitate bună	3
	Calitate excelentă cu un interval mare de markeri relevanți pentru validare convergentă și divergentă	4
10.1.12	<b>Cât de vechi sunt studiile de validitate?</b>	
	Este dificil de formulat o regulă generală pentru luarea în considerare a "vârstei" cercetării. Pentru testele care intenționează să măsoare constructele într-o zonă în care s-au produs evoluții teoretice importante, cercetarea veche de 15 ani poate fi aproape inutilă, în timp ce pentru alte teste, cercetarea veche de 20 ani (sau chiar mai veche) poate fi relevantă.	
	Număr de ani	.....
10.1.13	<b>Validitatea constructului - adecvarea generală</b>	
	Această evaluare globală este obținută prin utilizarea evaluărilor de la punctele 10.1.1 - 10.1.12. <i>Nu realizați pur și simplu media numerelor pentru a obține un rating global.</i>	
	În plus față de rezultatele cercetării validității constructului, pentru evaluarea finală ar trebui să țineți cont și de utilizarea corectă a tehnicilor de analiză (de exemplu, nivelul de semnificație corectat pentru corelarea instrumentului cu alte instrumente fără ipoteze clare, așa-numitul "fishing"), indiferent dacă eșantioanele de cercetare sunt similare cu grupul (grupurile) pentru care se intenționează testul (de exemplu, o eterogenitate mai mare va "umfla" corelațiile, eșantioanele de studenți pot da rezultate care nu pot fi generalizate), dimensiunea eșantionului (oanelor) de cercetare, calitatea altor instrumente care sunt utilizate (de exemplu, în studii de validitate convergentă și discriminatorie), și vârsta studiilor.	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2

	Bun	3
	Excelent	4

## 10.2. Validitatea relativă la criteriu

Dovezile de validitate relativă la criteriu (validitate concurrentă și predictivă) se referă la studii în care măsurile criteriilor din lumea reală (de ex., nu alte scoruri ale instrumentului) au fost corelate cu scalele. Studiile predictive se referă, în general, la situațiile în care evaluarea a fost efectuată într-un punct "calitativ" diferit în timp, față de măsurarea criteriului – ex., pentru o măsură de selecție legată de muncă, destinată să prezică succesul la locul de muncă, instrumentul ar fi fost realizat la momentul selecției - mai degrabă decât să fie o chestiune legată de intervalul de timp dintre măsurarea instrumentului și a criteriului. Studiile pot fi, de asemenea, "postdictive", de exemplu, în cazul în care scorurile la un test de selecție potențială sunt corelate cu evaluările managerilor de rang anterior în ceea ce privește performanța. Practic, dovezile privind validitatea criteriilor sunt necesare pentru toate tipurile de teste. Cu toate acestea, atunci când se menționează în mod explicit în manual faptul că utilizarea testului nu servește scopurilor de predicție (cum ar fi testele educaționale care măsoară progresul), validitatea criteriului poate fi considerată "neaplicabilă".

<b>10.2</b>	<b>Validitate relativă la criteriu</b>	
10.2.1	Tip(uri) de studiu(ii) ale criteriului ( <i>selectați atâtea câte se aplică</i> )	
	Predictiv	[ ]
	Concurent	[ ]
	Postdictiv	[ ]
10.2.2	Mărimile eșantioanelor	
	Neaplicabil	0
	Nicio informație oferită	1
	Un studiu inadecvat (ex., eșantion mai mic de 100)	2
	Un studiu adecvat (ex., eșantion de 100-200)	3
	Un studiu mare (ex., eșantion mai mare de 200) sau mai mult decât adecvat ca mărime	4

10.2.3	Calitatea măsurătorilor criteriului	
	Nicio informație oferită	0
	Calitate inadecvată	1
	Calitate adecvată	2
	Calitate bună	3
	Calitate excelentă cu privire la fidelitatea și reprezentativitatea constructului criteriului	4
10.2.4	<p>Puterea relației dintre test și criterii</p> <p>Este dificil să se stabilească criterii clare pentru evaluarea mărimii coeficienților de validitate a criteriului unui instrument. O validitate a criteriului de 0,20 poate avea o utilitate considerabilă în unele situații, în timp ce unul din 0,40 poate avea o valoare mică în altele. Un coeficient de 0,30 poate fi considerat bun în selecția personalului, în timp ce în situații educaționale, coeficienții mai mari sunt mai obișnuiți. Din aceste motive, evaluările ar trebui să se bazeze pe judecata și expertiza evaluatorilor, și nu pur și simplu derivate prin medierea seturilor de coeficienți de corelare. Instrucțiunile oferite sunt bazate pe Hemphill (2003, vezi și Meyer et al., 2001) și privesc corelațiile care nu sunt corectate pentru atenuarea predictorului sau a criteriului. Cu toate acestea, coeficienții pot fi corecți pentru restricționarea intervalului.</p>	
	<p>Intervalele de mai jos se referă la coeficienții de validitate, deoarece <i>corelațiile</i> dintre teste și criterii reprezintă modalitatea cea mai utilizată de a reprezenta validitatea criteriului. Cu toate acestea, în special pentru utilizarea în situații clinice, datele privind sensibilitatea și specificitatea unui test pot oferi informații mai utile cu privire la relația dintre un test și un criteriu. Curbele ROC reprezintă o metodă populară de cuantificare a sensibilității și a specificității. Swets (1988) prezintă o vedere de ansamblu a valorilor curbelor ROC în diferite zone. Pentru anumite tipuri de diagnostic medical, valorile se situează între 0,81 și 0,97, pentru detectarea minciunii, între 0,70 și 0,95 și pentru realizarea educațională (pass / fail) între 0,71 și 0,94. Aceste valori pot fi utilizate ca recomandări, dar este lăsat în seama expertizei evaluatorului să decidă în ce măsură testul poate aduce o contribuție utilă la decizia respectivă. Acest aspect se aplică, de asemenea, și când se raportează și alți indicatori, cum ar fi valoarea predictivă pozitivă și negativă a unui test, rata probabilității etc.</p>	
	Nicio informație oferită	0
	Inadecvat ( $r < 0.20$ )	1
	Adecat ( $0.20 \leq r < 0.35$ )	2
	Bun ( $0.35 \leq r < 0.50$ )	3

	Excelent ( $r \geq 0.50$ )	4
10.2.5	<p>Cât de vechi sunt studiile de validitate?  Este dificil de formulat o regulă generală pentru luarea în considerare a "vârstei" cercetării. Pentru testele care intenționează să prezică comportamente într-un mediu cu schimbări rapide, cercetarea veche de 15 ani poate fi aproape inutilă, în timp ce pentru alte teste, cercetarea veche de 20 ani (sau chiar mai veche) poate fi relevantă.</p>	
	Număr de ani	.....
10.2.6	<p><b>Validitatea relativă la criteriu – adecvarea generală</b>  Această evaluare generală este obținută pe baza evaluărilor de la punctele 10.2.1 - 10.2.5. <i>Nu realizați pur și simplu media numerelor pentru a obține un rating global.</i>  În afară de rezultatele cercetării validității criteriului, pentru evaluarea finală se ține cont și de faptul dacă se utilizează procedurile potrivite și tehnicile de analiză (ex., există contaminare a criteriilor, corecția pentru atenuare, validarea încrucișată), dacă eșantioanele de cercetare sunt similare cu grupul (grupurile) pentru care se intenționează testul (de exemplu, corecția pentru restricționarea intervalului), mărimea eșantionului (eșantioanelor) de cercetare, calitatea instrumentelor de criterii utilizate (de exemplu, există o deficiență de criterii) vârsta studiilor.</p>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4

### 10.3. Validitatea generală

Atunci când se evaluează validitatea generală, este important să se țină seama de importanța acordată validității de construct ca fiind cel mai bun indicator dacă un test măsoară ceea ce pretinde să măsoare. În unele cazuri, principalele dovezi ale acestui fapt ar putea fi sub forma studiilor legate de criterii. Un astfel de test ar putea avea o evaluare "adecvată" sau mai bună pentru validitatea de criteriu și o evaluare mai puțin adecvată pentru validitatea de construct. În general, evaluarea validității globale va fi egală fie cu validitatea de construct, fie cu validitatea relativă la criteriu, oricare dintre acestea este mai mare. Cu toate acestea, în funcție de scopul testului, unul dintre aceste tipuri de dovezi poate fi considerat mai relevant decât celălalt. Evaluarea validității globale nu ar trebui să fie considerată ca o medie sau ca cel mai mic numitor comun.

10.3	<b>Validitate – adecvare generală</b>	
	Acest rating global este obținut prin utilizarea evaluărilor de la punctele 10.1.1 –10.2.6. <i>Nu realizați pur și simplu media numerelor pentru a obține un rating global.</i>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
Excelent	4	

**Comentariile evaluatorilor privind validitatea** (toate dovezile de validitate incluse).  
Comentariile referitoare la generalizarea echivalenței / validității ar trebui, de asemenea, realizate aici (dacă este cazul).

## 11. Calitatea rapoartelor generate de computer

Evaluarea rapoartelor generate de calculator este dificilă, deoarece mulți furnizori vor dori, în mod lesne de înțeles, să-și protejeze proprietatea intelectuală în algoritmi și regulile de notare. În practică, ar trebui să existe suficientă informație disponibilă în manualul tehnic pentru evaluare, care să descrie evoluția procesului de raportare și raționamentul său, și prin rularea unui eșantion de cazuri de testare a configurațiilor de scoruri. În mod ideal, documentația ar trebui să descrie, de asemenea, procedurile care au fost utilizate pentru a testa generarea de rapoarte pentru exactitate, coerență și relevanță. În scopul examinării, cel puțin trei rapoarte bazate pe profile cu scoruri diferite, inclusiv scorurile efective, ar trebui să fie furnizate, chiar dacă algoritmi pentru generarea rapoartelor sunt confidențiale.

Itemii vor fi clasificați ca n/a sau de la 0 la 4, sunt oferite "pragurile" pentru o evaluare "excelentă" (4).											
11.1	<p><b>Scop sau domeniu de aplicare</b></p> <p>Rapoartele pot fi văzute ca variind, atât în acoperire, cât și în specificitate. Rapoartele pot varia, de asemenea, în ceea ce privește tipurile de persoane pentru care sunt potrivite. În unele cazuri, este posibil să fie furnizate rapoarte individuale adaptate pentru diferite grupuri de destinatari.</p> <ul style="list-style-type: none"> <li>• <i>Raportul acoperă intervalul atributelor măsurate de instrument?</i></li> <li>• <i>Face acest lucru la un nivel de specificitate justificabil din punct de vedere al nivelului de detaliere obținut din punctajul instrumentului?</i></li> <li>• <i>Poate fi justificată "granularitatea" raportului (ex., numărul de benzi de scorare distincte pe o scală, care sunt utilizate pentru reprezentări grafice pe diferite unități de text utilizate în raport) în ceea ce privește erorile de măsurare a scalelor?</i></li> <li>• <i>Raportul este conceput pentru aceleași populații de oameni pentru care instrumentul a fost dezvoltat? (de exemplu grupuri pentru care grupurile de norme sunt relevante sau pentru care există date relevante privind criteriile etc.).</i></li> </ul>										
	<table border="1"> <tr> <td>Nicio informație oferită</td> <td>0</td> </tr> <tr> <td>Inadecvat</td> <td>1</td> </tr> <tr> <td>Adecvat</td> <td>2</td> </tr> <tr> <td>Bun</td> <td>3</td> </tr> <tr> <td>Excelentă: potrivire excelentă între domeniul de aplicare al instrumentului și domeniul de aplicare al raportului, nivelul de specificitate din raport fiind egal cu</td> <td>4</td> </tr> </table>	Nicio informație oferită	0	Inadecvat	1	Adecvat	2	Bun	3	Excelentă: potrivire excelentă între domeniul de aplicare al instrumentului și domeniul de aplicare al raportului, nivelul de specificitate din raport fiind egal cu	4
Nicio informație oferită	0										
Inadecvat	1										
Adecvat	2										
Bun	3										
Excelentă: potrivire excelentă între domeniul de aplicare al instrumentului și domeniul de aplicare al raportului, nivelul de specificitate din raport fiind egal cu	4										

	nivelul de detaliere măsurat de scale. Buna utilizare a tuturor scorurilor raportate de instrument.	
11.2	<p><b>Fidelitate</b></p> <ul style="list-style-type: none"> <li>• <i>Cât de consecvente sunt rapoartele în interpretarea lor de seturi similare de scoruri?</i></li> <li>• <i>Dacă conținutul rapoartelor este variat (ex., prin selectarea aleatorie a unităților de text echivalente), aceasta se face într-o manieră satisfăcătoare?</i></li> <li>• <i>Sunt interpretarea scorurilor și diferențele dintre scoruri justificabile în ceea ce privește erorile de măsurare a scalei?</i></li> </ul>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelentă: Consistență excelentă în interpretare și avertismente corespunzătoare oferite pentru afirmații, interpretare și recomandări privind erorile de măsurare care stau la baza acestora.	4
11.3	<p><b>Relevanță sau validitate</b></p> <p>Legătura dintre instrument și conținutul raportului poate fi explicată fie în cadrul raportului, fie documentată separat. În cazul în care rapoartele se bazează pe o evaluare clinică, procesul prin care expertul (experții) a/au produs conținutul și regulile care raportează scorurile la conținut trebuie să fie documentate.</p> <ul style="list-style-type: none"> <li>• <i>Cât de puternică este relația dintre conținutul raportului și scorurile de pe instrument? În ce măsură explică raportul sau deviază de la informațiile furnizate de scorurile instrumentului?</i></li> <li>• <i>Conținutul raportului relaționează în mod clar cu caracteristicile măsurate de instrument?</i></li> <li>• <i>Oferă concluzii rezonabile cu privire la criteriile cu care ne putem aștepta ca astfel de caracteristici să fie corelate?</i></li> <li>• <i>Ce dovezi empirice sunt prezentate pentru a arăta că aceste relații există de fapt?</i></li> </ul> <p>Este relevant să se ia în considerare atât validitatea de construct a unui raport (ex., măsura în care acesta oferă o interpretare conformă cu definiția constructelor subiacente), cât și validitatea de criteriu (ex., în cazul în care se fac aserțiuni care pot fi raportate la date empirice).</p>	
	Nicio informație oferită	0
	Inadecvat	1



	Adecvat	2
	Bun	3
	Excelent: Relația dintre scale și conținutul raportului, cu justificări clare furnizate.	4
11.4	<b>Corectitudine, sau libertate față de biasul sistematic</b> <ul style="list-style-type: none"> <li>• <i>Este posibil ca conținutul raportului și al limbii folosite să creeze impresii de inadecvare pentru anumite grupuri?</i></li> <li>• <i>Raportul clarifică toate domeniile de posibile biasuri în rezultatele instrumentului?</i></li> <li>• <i>Sunt disponibile formulare în limbi alternative? Dacă da, au fost luate măsuri adecvate pentru a asigura echivalența lor?</i></li> </ul>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent: Avertismente clare și explicații privind posibile biasuri, disponibile în toate limbile relevante ale utilizatorilor.	4
11.5	<b>Acceptabilitate</b> <p>Acest lucru depinde în mod substanțial de complexitatea limbajului utilizat în raport, de complexitatea construcțiilor descrise și de destinația scopului său.</p> <ul style="list-style-type: none"> <li>• <i>Este rezonabil ca forma și conținutul raportului să fie acceptabile pentru destinatarii vizați?</i></li> <li>• <i>Este raportul redactat într-o limbă potrivită pentru nivelurile de aritmetică și alfabetizare a cititorului țintit?</i></li> </ul>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent: acceptabilitate foarte ridicată, bine conceput și potrivit publicului țintă.	4

11.6	<b>Lungime</b>	
	Acesta este un aspect pragmatic și ar trebui să se reflecte în evaluare. Rapoartele prea lungi pot fi un indicator al supra-interpretării scorurilor. În general, rapoartele care iau în medie mai mult de o pagină per scală (cu excepția paginilor de titlu, a anunțurilor privind drepturile de autor etc.), pot fi prea lungi și supra-interpretate.	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4
11.7	<b>Adecvarea generală a rapoartelor generate de computer</b>	
	Acest rating global este obținut prin utilizarea evaluărilor de la punctele 11.1 -11.6. <i>Nu realizați pur și simplu media numerelor pentru a obține un rating global.</i>	
	Nicio informație oferită	0
	Inadecvat	1
	Adecvat	2
	Bun	3
	Excelent	4

### **Comentariile evaluatorilor asupra rapoartelor generate de calculator**

Evaluarea poate lua în considerare aspecte suplimentare, cum ar fi dacă rapoartele conțin informații privind verificarea consecvenței răspunsurilor, măsurilor de bias la raspuns (ex., măsurile de tendință centrală în evaluare ) și alți indicatori ai încrederii cu care scorurile persoanei pot fi interpretate. Pot fi incluse comentarii privind complexitatea algoritmilor, de ex., dacă sunt luate în considerare simultan mai multe scale, cum sunt tratate profilurile scalei, etc. O astfel de complexitate ar trebui, bineînțeles, să fie susținută de un raționament clar în manual.

## 12. Evaluarea finală și tipuri de rezoluții

### Raportul evaluativ al testului

Această secțiune ar trebui să conțină o concluzie concisă și clar argumentată cu privire la test. Ar trebui să descrie argumentele sale pro și contra și să dea câteva recomandări generale cu privire la modul și momentul în care ar putea fi utilizate - împreună cu avertismentele (când este cazul) despre momentul când nu ar trebui utilizate.

Un rezumat al tuturor punctelor pozitive sau negative în legătură cu testele adaptate și traduse ar trebui inclus aici. În anexă este prezentat un ghid de tip check-list.

Evaluarea ar trebui să acopere aspecte precum adecvarea instrumentului pentru diferite funcții de evaluare sau domenii de aplicare; orice necesități speciale de formare sau competențe speciale necesare; dacă cerințele de formare sunt stabilite la nivelul corespunzător; ușurință în utilizare; calitatea și cantitatea informațiilor furnizate de furnizor și dacă există informații importante care nu sunt furnizate utilizatorilor și în care există probleme care rezultă din traducerea sau adaptarea instrumentului (vezi Anexa).

Includeți observații cu privire la orice cercetare despre care se știe că este în curs de desfășurare și despre planurile furnizorilor pentru viitoarele dezvoltări și perfecționări, etc.

### Concluzii

<p><b>Recomandări (alegeți una)</b></p> <p>Recomandarea relevantă din listă trebuie indicată. În mod normal, acest lucru va necesita un comentariu, o justificare sau o calificare. Ar trebui adăugată o scurtă declarație referitoare la situațiile și modurile în care instrumentul ar putea fi utilizat și la avertismentele cu privire la posibilele zone de utilizare incorectă.</p> <p><b>Toate caracteristicile listate mai jos trebuie să primească o notă, n/a, 2, 3, sau 4, dacă un instrument urmează să fie "recomandat" pentru uz general (căsuțele 4 sau 5).</b></p> <p>8 Norme</p> <p>9 Fidelitate-general</p> <p>10 Validitate-general</p> <p>11 Rapoarte generate de calculator</p> <p>Dacă oricare dintre aceste note sunt 0 sau 1, instrumentul va fi clasificat ca Recomandare 1, 2, sau 3, sau va fi clasificat ca "Altul", împreună cu o explicație satisfăcătoare.</p>	1 Necesită dezvoltare ulterioară. Potrivit numai pentru utilizare în cercetare, nu pentru utilizare în practică	[ ]
	2 Potrivit doar pentru utilizarea de către un utilizator expert (depășind nivelul de calificare a utilizatorului EFPA 2), în condiții controlate cu atenție sau în domenii foarte limitate de aplicare	[ ]
	3 Potrivit pentru utilizare sub supraveghere în zona (domeniile) de aplicare definită/e de distribuitor, de către orice utilizator cu o competență generală în utilizarea testelor și administrarea testelor (depășind nivelul de calificare a utilizatorului EFPA 2)	[ ]
	4 Pot fi utilizate în zona (domeniile) de aplicare definită/e de distribuitor, de către utilizatorii de test care îndeplinesc cerințele de calificare specifice ale distribuitorului (cel puțin nivelul EFPA pentru calificarea utilizatorului 2)	[ ]
	5 Potrivit pentru auto-evaluare nesupravegheată în zona (domeniile) de aplicare definită/e de distribuitor	[ ]
	6 Altul .....	[ ]

În urma analizei / evaluării efectuate de către membrii comisiei, Comisia Metodologică poate propune Comitetului Director acordarea a **trei tipuri de rezoluții**:

- a) aviz pe perioadă nedeterminată
- b) amânare (cu *minor revision* și acordare de termen)
- c) respins

Comisia Metodologică va întocmi o motivare prin care să justifice tipul de rezoluție, dar și recomandări pentru îmbunătățirea calităților psihometrice ale probelor. De asemenea, vor fi aprobate și eliberate instrucțiunile de utilizare ale metodei sau tehnicii de evaluare sau asistență psihologică.

*Avizul eliberat pe perioadă nedeterminată* corespunde unei metode sau tehnici de evaluare sau asistență psihologică care îndeplinește toate condițiile de calitate necesare utilizării de către psihologii cu drept de liberă practică.

*Amânarea cu minor revision* corespunde unei metode sau tehnici de evaluare sau asistență psihologică care îndeplinește condițiile minime de calitate necesare utilizării de către psihologii cu drept de liberă practică, însă necesită completări și/sau dovezi suplimentare. În aceste condiții, proba va fi reanalizată de către comisie, conform termenului acordat (maxim 6 luni). Reanalizarea probei presupune redopunerea dosarului și achitarea unui procent de 50% din contravaloarea taxei inițiale. O probă poate beneficia de o singură sesiune/procedură de reanalizare a dosarului.

*Respingerea cererii de acordare a avizului* pentru metoda sau tehnica de evaluare sau asistență psihologică corespunde neîndeplinirii condițiilor minime de calitate necesare utilizării de către psihologii cu drept de liberă practică sau prezentării unor date incomplete.

Avizul eliberat de către Comitetul director se rezumă la aria de aplicabilitate specificată în caracteristicile domeniului vizat, și nu poate fi extins spre alte domenii de activitate fără o avizare prealabilă.

### **Standarde minimale ale rezoluțiilor, conform cerințelor prevăzute în tabel**

**Avizarea pe perioadă nedeterminată** (recomandare pentru uz general, casetele 4 și 5) - Note de 2, 3 sau 4 pentru patru criterii majore ale testului: norme, fidelitate, validitate, rapoarte generate de calculator.

**Amânare (cu *minor revision*)** - Note de 1 sau 2, iar instrumentul va primi recomandări potrivit casetelor 1, 2 sau va fi încadrat în categoria "Altul".

**Respins** – Note de 0 sau 1, iar instrumentul va primi recomandări potrivit casetelor 2, 3 sau va fi încadrat în categoria "Altul".

## PARTEA A TREIA – AVIZAREA METODELOR DE ASISTENȚĂ / INTERVENȚIE PSIHOLOGICĂ

Comisia de Metodologie elaborează cadrul general de evaluare și avizare a Metodelor de asistență psihologică.

Scopul principal al avizării metodelor de intervenție sau tratamentelor psihologice este de a stabili baza științifică a unei metode. Astfel, validitatea științifică indică dacă o modalitate/metodă de intervenție este conformă cu o teorie unanim recunoscută și dacă sunt aplicate metode științifice în stabilirea eficienței acesteia. În cazul în care o metodă nu realizează acest lucru, atunci aceasta acționează mai degrabă ca un *sistem de credințe*.

*Dovadă* este acea informație oferită ca răspuns în forma unui raport sau a unei documentații. Sunt considerate dovezi: documentele emise de experții în domeniu, cercetările și studiile publicate, studiile clinice. Nu sunt considerate dovezi: opinii personale, puncte de vedere sau credințe.

Evaluatorii analizează dovezile (afirmate, documentate și depuse la dosar) și decid relevanța acestora, potrivit criteriilor de adecvare, suficiență, caracter științific.

Evaluarea metodei se realizează prin analiza dovezilor științifice invocate. În general, dovezile științifice ale metodei / tratamentului trebuie să răspundă mai multor criterii.

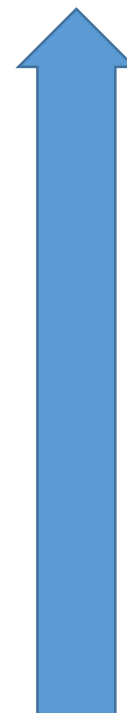
În vederea alinierii la modelul EBP (Evidence-Based-Practice), sau practică validată științific, Comisia Metodologică utilizează **Grila de evaluare a calității dovezilor științifice, potrivit sistemului GRADE** (Atkins et al., 2004; Balshemet et al., 2011; Guyatt et al., 2011):

- (a) designul studiului (care apare și în piramida calității dovezilor științifice)
- (b) calitatea studiilor
- (c) consistența rezultatelor
- (d) similaritatea studiilor cu situația practică de rezolvat.

Piramida încrederii privind eficiența unei intervenții psihologice în funcție de calitatea metodologică a studiilor pe care se bazează:

*Dovezi științifice solide*

1. Metaanalize pe studii primare experimentale (eventual cvasi-experimentale)
2. Studii randomizate privind mecanismele schimbării (explicative)
3. Studii experimentale (randomizate) cauzal-descriptive (efficacy)
4. Studii experimentale (randomizate) pragmatice (effectiveness)
5. Studii cvasiexperimentale
6. Metaanalize pe studii primare non-experimentale
7. Studii non-experimentale longitudinale
8. Studii non-experimentale transversale
9. Studii de caz
10. Opiniile unor experți



Cu cât dovezile se situează mai spre partea superioară a piramidei, cu atât ele sunt mai solide.

În baza datelor referitoare la design, la calitatea studiilor, la robustețea și omogenitatea efectelor și la gradul de aplicabilitate a acestora la situația practică întâlnită, sistemul GRADE clasifică intervențiile în patru categorii diferite, din perspectiva gradului de încredere că acestea vor funcționa pentru situația practică întâlnită.

**Niveluri de încredere în reușita intervenției (punctaj între paranteze):**

- **ridicat (4)** – probabilitate scăzută ca studii viitoare să conducă la modificarea încrederii în eficiența intervenției;
- **moderat (3)** – studii viitoare pot avea un impact relevant pentru concluziile referitoare la eficiența intervenției;
- **scăzut (2)** – probabilitate ridicată ca studiile viitoare să aibă impact important asupra concluziilor, fiind posibil să asistăm la situația inversării direcției efectului;
- **foarte scăzut (1)** – orice direcție a efectului este posibilă.



### Scoruri de pornire inițială:

- dovezi obținute în baza unor studii experimentale – 4;
- dovezi obținute în baza unor studii non experimentale (observaționale) sau cvasiexperimentale – 2;
- dovezi bazate pe opinii – 1.

Aceste punctaje sunt ajustate, după cum urmează:

### Scăderi de punctaj:

- –1 sau –2 puncte pentru calitatea scăzută a studiilor sub standardul implicat de designul respectiv;
- –1 pentru număr mic de studii sau pentru cazul în care se observă efecte eterogene (variații mari);
- –1 sau –2 puncte penalizare pentru studiile care au puține similitudini cu grupul țintă vizat în practică;
- –1 pentru situații de prezentare trunchiată a rezultatelor (reporting bias).

### Creșteri de punctaj:

- +1 dacă variabilele străine necontrolate ar fi putut afecta datele doar în direcția reducerii efectului;
- +1 dacă este manipulată VI (studii cvasiexperimentale);
- +1 sau +2 dacă există asocieri puternice între variabile (risc relativ dublu, Cohen d cu aproximație de 0,40), respectiv foarte puternice (risc relativ de 5 ori mai mare/mai mic, Cohen d aproximativ 0,90).

**Notă:** Întrucât metodele sunt specifice unei anumite populații sau unei anumite problematici, acestea nu pot fi avizate la modul general.

**Exemplificare:** Metoda X este eficientă în zona anxios-depresivă, însă nu neapărat în zona de trăsături schizotipale accentuate. / Dacă Metoda Y se adresează unei populații subclinice, ea nu va fi în mod obligatoriu eficientă pentru populația clinică sau nonclinică.

Metoda nu poate funcționa în contexte în care nu și-a dovedit eficiența, ceea ce înseamnă că avizul se va solicita, în cazul fiecărei metode, pentru o anumită utilitate.

Prin urmare, avizul va cuprinde sintagma ”**Metoda X – cu aplicație în zona Y**”. Pentru conformitatea Metodă – Aplicabilitate, vor fi consultate ghidurile internaționale în domeniu (NICE, APA, etc.). / solicitantul va aduce dovezi de conformitate, conform prevederilor internaționale.

### **Repere pentru întocmirea dosarului**

În vederea analizei, vor fi prezentate dovezile științifice ale metodei, după cum urmează:

- Fundamentare teoretică solidă – teoriile / modelele teoretice ale metodei
- Recunoaștere generală (națională și internațională) în domeniu
- Literatură de specialitate descriptivă (manuale, tratate, etc.) – cel mult 100 de titluri
- Rezultate publicate în jurnale cu peer-review
- Dovezile de absență a lezării și prejudicierii
- Studii conform piramidei calității dovezilor științifice
- Dovezi de fidelitate și validitate a rezultatelor / eficienței metodei
- Numărul de ore de formare în domeniu / în metodă, specific fiecărei metode
- Posibilitatea utilizării metodei într-un context comun

### **Dispoziții finale**

1. Comisia de Metodologie poate solicita aviz consultativ comisiilor de specialitate.
2. Conținutul dosarului pentru solicitanți va fi detaliat în termen de 90 de zile de la data adoptării.
3. Lista actualizată a metodelor și tehnicilor de evaluare și asistență psihologică va fi afișată pe site.

## Referințe bibliografice

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Baker, F. B., & Kim, S. H. (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.
- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment, 12*, 62–71.
- Bartram, D. (2002a). *EFPA Review Model for the description and evaluation of psychological instruments: Version 3.2. Evaluation Form*. Brussels: EFPA Standing Committee on Tests and Testing (September, 2002).
- Bartram, D. (2002b). *EFPA Review Model for the description and evaluation of psychological instruments: Version 3.2. Notes for Reviewers*. Brussels: EFPA Standing Committee on Tests and Testing (September, 2002).
- Bartram, D., & Hambleton, R. K. (Eds.) (2006). *Computer-based testing and the Internet*. Chichester, UK: Wiley and Sons.
- Bartram, D., Lindley, P. A., & Foster, J. M. (1990). *A review of psychometric tests for assessment in vocational training*. Sheffield, UK: The Training Agency.
- Bartram, D., Lindley, P. A., & Foster, J. M. (1992). *Review of psychometric tests for assessment in vocational training*. BPS Books: Leicester.
- Bechger, T., Hemker, B., & Maris, G. (2009). *Over het gebruik van continue normering* [On the use of continuous norming]. Arnhem, The Netherlands: Cito.
- Bennett, R. E. (2006). Inexorable and inevitable: The continuing story of technology and assessment. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the Internet* (pp. 201-217). Chichester, UK: Wiley and Sons.
- Brennan, R. L. (Ed.) (2006). *Educational measurement*. Westport, CT: ACE/Praeger.
- Cohen, A. S., & Kim, S. H. (1998). An investigation of linking methods under the graded response model. *Applied Psychological Measurement, 22*, 116-130.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (3<sup>rd</sup> ed.). New York: Harper & Row.
- Downing, S. M., & Haladyna, T. M. (Eds.) (2006). *Handbook of test development*. Hillsdale, NJ: Erlbaum.
- Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and testing. In R. L. Brennan (Ed.), *Educational measurement* (pp. 471-515). Westport, CT: ACE/Praeger.
- Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie. Inleiding in de theorie van de psychologische test en zijn toepassingen* (4e herziene druk) [Test theory. Introduction in the theory and application of psychological tests (4th revised ed.)]. Houten, The Netherlands: Bohn Stafleu van Loghum.
- Embretson, S. E. (Ed.) (2010). *Measuring psychological constructs. Advances in model-based approaches*. Washington, D. C.: American Psychological Association.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Evers, A. (2001a). Improving test quality in the Netherlands: Results of 18 years of test ratings. *International Journal of Testing, 1*, 137–153.

- Evers, A. (2001b). The revised Dutch rating system for test quality. *International Journal of Testing*, 1, 155–182.
- Evers, A., Braak, M., Frima, R., & van Vliet-Mulder, J. C. (2009-2012). *Documentatie van Tests en Testresearch in Nederland* [Documentation of Tests and Testresearch in The Netherlands]. Amsterdam: Boom test uitgevers.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de Kwaliteit van Tests (geheel herziene versie; gewijzigde herdruk)* [COTAN Rating system for test quality (completely revised edition; revised reprint)]. Amsterdam: NIP.
- Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., et al. (2012). Testing practices in the 21<sup>st</sup> Century: Developments and European psychologists' opinions. *European Psychologist*, in press.
- Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure and results. *International Journal of Testing*, 10, 295-317.
- Khan, K.S., et al. (2001). *CRD Report 4. Stage II. Conducting the review. phase 5. Study quality assessment*. York, UK: Centre for Reviews and Dissemination, University of York.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-334.
- Hambleton, R. K., Jaeger, R. M., Plake, B. S., & Mills, C. (2000). Setting performance standards on complex educational assessments. *Applied Psychological Measurement*, 24, 355–366.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.
- Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58, 78-80.
- International Test Commission. (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*. Bruxelles, Belgium: Author.
- Kersting, M. (2008). DIN Screen, Version 2. Leitfaden zur Kontrolle und Optimierung der Qualität von Verfahren und deren Einsatz bei beruflichen Eignungsbeurteilungen [DIN Screen, Version 2. Guide line for monitoring and optimizing the quality of instruments and their application in proficiency assessment procedures.]. In M. Kersting. *Qualitätssicherung in der Diagnostik und Personalauswahl - der DIN Ansatz* (S. 141-210) [Guaranteeing quality in diagnostics and personnel selection (p. 141-210)]. Göttingen: Hogrefe.
- Lindley, P. A. (2009). *Reviewing translated and adapted tests: Notes and checklist for reviewers:5 May 2009*. Leicester, UK: British Psychological Society. Retrieved from. eu/professional-development/tests-and-testing.
- Lindley, P.A. (2009, July). Using EFPA Criteria as a common standard to review tests and instruments in different countries. In D.Bartram (Chair), *National approaches to test quality assurance*. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.
- Lindley, P., Bartram, D., & Kennedy, N. (2004). *EFPA Review Model for the description and evaluation of psychological tests: test review form and notes for reviewers: Version 3.3*. Leicester, UK: British Psychological Society (November, 2004).
- Lindley, P., Bartram, D., & Kennedy, N. (2005). *EFPA Review Model for the description and evaluation of psychological tests: test review form and notes for reviewers: Version 3.41*. Brussels: EFPA Standing Committee on Tests and Testing (August, 2005).

- Lindley, P., Bartram, D., & Kennedy, N. (2008). *EFPA Review Model for the description and evaluation of psychological tests: test review form and notes for reviewers: Version 3.42*. Brussels: EFPA Standing Committee on Tests and Testing (September, 2008).
- Lindley, P. A. (Senior Editor), Cooper, J., Robertson, I., Smith, M., & Waters, S. (Consulting Editors). (2001). *Review of personality assessment instruments (Level B) for use in occupational settings. 2<sup>nd</sup> Edition*. Leicester, UK: BPS Books.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*, 128-165.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton.
- Moosbrugger, H., Kelava, A., Hagemeister, C., Kersting, M., Lang, F., Reimann, G., et al. (2009, July). The German Test Review System (TBS-TK) and first experiences. In D. Bartram (Chair), *National approaches to test quality assurance*. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.
- Moreno, R., Martínez, R. J., & Muñoz, J. (2006). New guidelines for developing multiple-choice items. *Methodology*, *2*, 65-72.
- Muñoz, J., & Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, *12*, 206-219.
- National Registry of Evidence-Based Practices and Programs (2007). *NREPP Review Criteria*.
- Nielsen, S. L. (2009, July). Test certification through DNV in Norway. In D. Bartram (Chair), *National approaches to test quality assurance*. Symposium conducted at The 11th European Congress of Psychology, Oslo, Norway.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3<sup>rd</sup> ed.). New York: McGraw-Hill.
- Parshall, C. G., Spray, J. A., Davey, T., & Kalohn, J. (2001). *Practical Considerations in Computer-based Testing*. New York: Springer Verlag.
- Prieto, G., & Muñoz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España [A model for the evaluation of test quality in Spain]. *Papeles del Psicólogo*, *77*, 65-71.
- Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Measurement*, *84*, 228-238.
- Tideman, E. (2007). Psychological tests and testing in Sweden. *Testing International*, *17*(June), 5-7.
- Sava, A. (2013). *Psihologia validata stiintific. Ghid practic de cercetare*. Iași: Polirom.
- 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*, *10*, 75-129.
- Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, *7*, 301-317.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285-1293.
- Testkuratorium. (2009). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 09. September 2009 [TBS-TK. Test review system of the board of testing of the Federation of German psychologists' associations]. *Report Psychologie*, *34*, 470-478.
- Van de Vijver, F. J. R., & Poortinga, Y. H. (2005). Conceptual and methodological issues in adapting tests. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Erlbaum.

- Van der Linden, W. J., & Glas, C. A. W. (Eds.) (2010). *Elements of adaptive testing*. London: Springer.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Ziegler, M., MacCann, C., & Roberts, R. (Eds.) (2011). *New perspectives on faking in personality assessment*. Oxford, UK: Oxford University Press.
- <http://www.gradeworkinggroup.org/>
- <https://www.europsychology.org/>

## Anexă

### Ghid al punctelor critice în cazul unui instrument tradus și / sau adaptat dintr-un context non-local

<b>Dezvoltare</b>	
Dovezi sau discuții despre	<i>Input de la vorbitori nativi de limbi noi</i>
	<i>Evaluarea mai multor experți în limbi și conținut (de testare)</i>
	<i>Traducerea din noua limbă în limba originală</i>
Proprietăți psihometrice de bază	<i>Performanța itemului</i>
	<i>Fidelitate</i>
<b>Norme</b>	
	<i>Este prevăzută o normă locală</i>
Normă non-locală	<i>Dovezi puternice de echivalență atât pentru versiunile de testare, cât și pentru eșantioane</i>
Norme internaționale	<i>Cerințe mai mari decât cele tipice ale probelor locale</i>
Natura eșantionului	<i>Echilibrarea surselor din eșantion</i>
	<i>Echivalența contextului diferitelor părți ale eșantionului</i>
Tipul măsurii	<i>Conținut verbal puțin sau deloc</i>
Echivalența versiunii de test	<i>Toate versiunile lingvistice sunt bine traduse / adaptate</i>
	<i>Unele grupuri au completat testul într-o limbă non-primară</i>
Asemănări ale scorurilor în diferite eșantioane	<i>În cazul în care există diferențe mari, acestea trebuie să fie luate în considerare și implicațiile în utilizare discutate</i>
Instrucțiuni privind generalizarea normelor	
<b>Echivalență/Fidelitate/Validitate</b>	
Invarianța în structura de construct	<i>Prin intermediul structurii factoriale, echivalența matricilor de corelare sau similaritatea modelelor de corelare cu măsuri standardizate</i>
Validitate similară a criteriului	<i>Cea mai puternică corelație cu competențe similare</i>
Modele similare de încărcare a scalei	<i>Elementele se corelează în același model cu alte scale</i>
	<i>Elementele de încărcare cele mai puternice / cele mai slabe sunt similare în limbile originale și noi</i>
Fidelitatea formei alternative	<i>Candidații bilingvi au profiluri similare în două limbi</i>
<b>Generalizarea validității</b>	
Generalizarea validității necesită dovezi puternice	<i>La traducerea testelor în familii lingvistice (ex., dintr-o limbă indo-europeană, într-o limbă semită)</i>
Generalizarea validității poate fi dedusă	<i>În cazul în care un test a fost tradus în mai multe limbi, o generalizare a validității poate fi dedusă din dovezile privind invarianța validității în traducerile anterioare: testul suedez a fost deja tradus în franceză, germană și italiană și s-a dovedit că are echivalență în aceste limbi.</i>