

## Principii metodologice în evaluarea și utilizarea probelor psihologice

Ioan Radu

Universitatea Babeș-Bolyai, Cluj-Napoca

Inventarul testelor psihologice, precum și ale altor mijloace de psihodiagnostic cuprinde în prezent instrumente de lucru de ordinul miilor. Totodată, numărul persoanelor care utilizează în activitatea lor profesională teste este destul de însemnat, formația lor psihologică fiind însă inegală, uneori redusă la minimum. De aici și riscul improvizăției, al aplicării incorecte a probelor, al utilizării neștiințifice, chiar a abuzului de teste în absența unei informații autentice despre valoarea și limitele fiecărui instrument de lucru.

De la început trebuie spus că probele psihologice nu pot fi alese doar după nume. Pentru a servi scopului pentru care a fost preferat, un test trebuie judecat în termeni de validitate, fidelitate, putere discriminativă și tabel de norme sau etalon (Anastasi, 1982). Alegerea testului, administrarea și cotarea sa, apoi interpretarea rezultatului reclamă o informație precisă despre creditul probei, despre eșantionul de etalonare, despre studiile de validare făcute etc. Toate aceste date se găsesc inserate, de regulă, în materialul de îndrumare sau manualul ce însoțește testul. Procurând deci un mijloc de psihodiagnostic, psihologul va avea grijă să obțină și materialul însoțitor, de asemenea, când se elaborează un test nou sau o variantă nouă va trebui să i se atașeze un material explicativ cât mai complet.

Termenul "test" este luat în sensul larg de procedură de măsurare sau evaluare cuprinzând:

- testele în sens strict (probe hârtie-creion, aparate cu destinație psihometrică și probe orale);
- inventare de personalitate, chestionare de interese, o bună parte din probele proiective și tehnicile clinice asociate (forme ale anamnezei);
- teste de cunoștințe.

Recomandările ce se formulează sunt de tip psihometric și se referă la instrumente de lucru destinate utilizării practice; ele nu privesc difuziunea în circuit închis, în scop de cercetare, a unor probe sau materiale. Desigur, normele se aplică în primul rând probelor standardizate dar ele sunt valabile în grade diferite pentru întreaga gamă de tehnici

de evaluare sau mijloace auxiliare aprecierii. Când ne referim la tehnicile proiective – unde datele extrase prezintă în primul rând un aspect calitativ – avem în vedere faptul că difuzarea lor în vederea utilizării practice impune cerințe unitare de validitate, fidelitate și interpretare. S-a remarcat de altfel tenta statistică a unor concluzii formulate de clinicieni când vorbesc de "fenomenul tipic", de "reacția obișnuită" a unei categorii de indivizi, de faptul că "majoritatea oamenilor privesc / cutare lucru / în felul..." etc. (cf. Standards, p.11).

Când se elaborează un test, un mijloc de diagnostic în general, se pornește de la o imagine a fenomenului vizat, care rezultă din datele cercetării psihologice sau este preluată din conștiința comună, cotidiană. Există, în arsenalul probelor psihologice, destule mijloace de psihodiagnostic născute din tatonări practice, oarecum pur empirice.

Se înțelege că nu se poate formula apriori o definiție completă a obiectului de studiu. Poate fi vorba doar de o definiție operațională, în limbajul datelor stăpânite în momentul respectiv, care se îmbogățește treptat prin aproximări succesive. În construcția probei se reflectă limitele conceptului de la care se pornește. Cu fiecare achiziție nouă în sfera cercetării, imaginea despre fenomenul studiat se completează, câștigă în profunzime și în consecință testul însuși devine susceptibil de amendare. Drept urmare, pe lângă tehnicitatea aplicării, practicianul trebuie să stăpânească teoria relativă la interpretarea unui test și să fie în măsură să aprecieze când ea este depășită. Spre exemplu, este anacronic să se utilizeze astăzi simple calcule aritmetice drept probe de aptitudini matematice; tot așa testele de aptitudine tehnică nu se mai pot reduce în prezent doar la mecanisme cu roți dințate.

Pornind de la considerațiile expuse, se recomandă:

1. Testele și manualele însoțitoare să facă obiectul unor revizii periodice, apreciindu-se că testele ale căror manuale nu au fost revizuite de mai bine de 15 ani să fie retrase din circulație sau în orice caz să fie preluate cu multă prudență, pe baza unui examen critic prealabil.

---

Necesitatea reviziei probelor de aptitudini reiese și din modificarea cerințelor posturilor de muncă, pentru care se face selecție, repartizare sau orientare.

2. Când un test este publicat într-o formă revizuită este necesar să se indice data noii ediții atât în testul propriu zis, cât și în manualul însoțitor. Cu aceeași ocazie trebuie să se precizeze natura și întinderea reviziei, precum și gradul de apropiere între forma veche și cea nouă, localizându-se concret în timp momentul recoltării noilor date și a stabilirii noilor tabele de norme. Așa cum se știe, orice modificare adusă probei, atât sub aspectul conținutului cât și al modului de administrare, atrage după sine schimbarea etaloanelor, precum și a tabelelor de conversie a cotelor brute în cote transformate.

3. Având în vedere tendința frecventă de a se întocmi forme prescurtate ale testelor existente pornind de la analiza de itemi și de la stabilirea zonei discriminative a testului, se preconizează următoarele: Când se alcătuieste o formă prescurtată a unei probe prin reducerea numărului de itemi sau prin detașarea unei părți a probei ca test separat, urmează să se obțină și să se difuzeze noile date privind testul prescurtat (indici de fidelitate și alte date). În plus, trebuie să se arate noile elemente de interpretare, precum și eroarea standard ce afectează datele obținute când se utilizează forma scurtă, în comparație cu proba inițială. (De exemplu, scala Terman – Merrill are și o formă prescurtată).

4. În legătură cu fiecare probă difuzată se va indica nivelul de calificare cerut de aplicarea și interpretarea datelor. Există teste și tehnici de evaluare pentru copii, respectiv pentru adulți care reclamă moduri diferențiate de administrare și interpretare; de asemenea, poate fi vorba de probe destinate mai ales populației generale, subiecților normali, fapt care necesită o formație psihologică obișnuită, după cum există instrumente de diagnostic – cum sunt testele proiective sau teste mintale individuale – care pe lângă formația psihologică avansată cer o experiență mai îndelungată desfășurată sub control competent. Spre exemplu, după unii autori pentru a interpreta cu relativă siguranță un protocol Rorschach sunt necesari aproximativ doi ani de antrenament; în același timp pentru un test de aptitudini este suficientă practica obișnuită prevăzută în anii studenției. “Când un test este recomandat pentru mai multe scopuri sau tipuri de inferență, manualul va trebui să indice volumul de antrenament cerut de fiecare

utilizare” (cf. Standards, p.16). Oricum, fiind vorba de interpretarea rezultatelor unei testări, oricâte indicații ar cuprinde manualul însoțitor, el nu poate suplini formația psihologică superioară oferită de facultățile de profil din universități.

5. Când în manualul însoțitor, se inserează studii de caz pentru ilustrarea modului de interpretare este necesar să se aleagă situații mai complicate, chiar puțin ambigue, care să se preteze la o discuție instructivă. De asemenea, dacă experiența scoate în relief riscul unei erori tipice (prin frecvența sa) în interpretarea datelor, manualul va avertiza cititorul cu toată claritatea. Spre exemplu, rezultatele la probele de inteligență generală sunt influențate de condiții socio-culturale defavorabile: în consecință, vor fi evitate afirmațiile sentențioase. Tot așa, se va avertiza psihologul practician asupra caracterului local al tabelului de norme.

6. Referitor la probele pentru care se prevede un timp limitat, trebuie să se precizeze ce influență are rapiditatea asupra rezultatelor obținute; se știe că o testare, pusă sub presiunea timpului, capătă uneori altă semnificație. Procedura este de a se compara cele două condiții: timp limitat și timp liber.

7. Când un test este destinat unei circulații mai restrânse, în scop de cercetare, se va menționa expres acest lucru în documentul atașat probei. La fel se va proceda cu testele de import, care trebuie supuse unei verificări prealabile, eventual modificate și adaptate (cu acordul autorului).

★

Să trecem în continuare la cerințe cu privire la *calitățile psihometrice* ale testelor.

Principala însușire a unui test o constituie validitatea sa, care ne indică – în sens larg – dacă proba măsoară / explorează cea ce își propune să măsoare și cât de bine (adecvat) realizează acest lucru. Orice testare psihologică se încadrează într-un obiectiv, o sarcină și urmărește să obțină un anumit gen de informație, de apreciere, care să servească unui tip de decizie. Indicii de validitate ne spun în ce măsură o probă răspunde unor funcții (sarcini) bine definite. Până la validare orice probă rămâne o simplă ipoteză.

În funcție de obiectivele testării se poate vorbi de cel puțin trei tipuri principale de validitate care se verifică și se completează reciproc. Este vorba de: (a) *validitate predictivă*, (b) *validitate de conținut*, și

(c) *validitate de construct* sau *ipotetico-deductivă*. La aceasta se adaugă anumite nuanțări: *validitate de aspect*, *validitate ecologică*, *factorială* etc.

a) *Validitatea predictivă* se referă îndeosebi la aspectul pragmatic al testării psihologice: în ce măsură informația furnizată de un examen pe bază de teste poate servi la formularea unei predicții cu privire la o situație viitoare, de exemplu cu privire la reușita școlară, reușita profesională, șansa unei deviații patologice etc. Astăzi se admite numai ideea *predicției pe termen scurt* și se contestă datelor psihometrice posibilitatea predicției pe termen lung.

Forma pe care o îmbracă determinarea acestui tip de validitate este, mai frecvent, aceea a indicilor de corelație dintre rezultatele la test și datele-criteriu. În practică, diferențierea realizată în funcție de criteriu este mai puțin nuanțată, adeseori datele respective se pretează doar la dihotomizare (admis/respins, bun/slab etc.) sau la trihotomizare (bun-mijlociu-slab). În consecință este indicată în asemenea cazuri utilizarea indicilor de corelație *r* bis și *r* tris. (Vezi Anexa A). Se cere ca valorile coeficienților de corelație să fie semnificative cel puțin la pragul de  $p = 0,05$ ; se consideră a fi de asemenea, utilă stabilirea ecuației de regresie (G. M. Smith 1971).

Paralel, validitatea este arătată prin indici – ca diferența între medii sau proporții – care scot în relief măsura în care testul separă grupele-criteriu. Validitatea predictivă mai poate fi prezentată sub forma tabelelor de previziune (“expectanță”) sau a diagramelor care redau grafic relația dintre test și criteriu.

În tabelele sau graficele de previziune se indică pe axa verticală datele la test, care sunt de obicei mai nuanțate (de exemplu stanine), iar pe orizontală se notează variabila-criteriu redusă uneori la două valori: succes, respectiv insucces profesional în procente (ca în Figura 1).

Practic, este vorba de un tabel de corelație în care se indică proporția de succes, respectiv de insucces (în profesii sau în școală), în funcție de datele-test.

O problemă dificilă este alegerea și definirea criteriului, care comportă el însuși mai multe laturi, față de care datele reținute drept criteriu formează o parte, un avantaj mai restrâns de situații, în cel mai bun caz un eșantion. Datele obținute cu privire la criteriu trebuie să satisfacă aceleași cerințe ca și testele, adică să reprezinte măsurări valide și fidele.

Normele stabilite pentru probele psihologice sunt în principiu valabile și pentru determinarea variabilei-criteriu. Se subliniază pe bună dreptate specificitatea criteriului și implicit a oricărui studiu de validitate. În acest sens este inexact să vorbim global de validitatea unei probe. Orice test este valid în raport cu un obiectiv specific și nu în orice situație și pentru orice grup de indivizi. Așadar, corelația semnificativă cu un criteriu extern nu pune punct final discuției despre validitatea unei probe, diversitatea și caracterul inedit al situațiilor lăsând mereu deschisă această problemă.

Pe de altă parte, studiile de validitate nu se rezumă, de obicei, la un singur criteriu global. Combinând datele în vederea obținerii unui singur criteriu compozit este important să se evite un amestec ambiguu, o reunire de informații necorelate.

Pornind de la aceste considerații se impun o seamă de norme:

1. Documentul de îndrumare atașat probei va trebui să indice în mod concret validitatea probei corespunzător fiecăreia din utilizările prevăzute sau în funcție de tipurile de decizie avute în vedere. De asemenea, se va specifica indicele de validitate în legătură cu fiecare criteriu semnalat. O ecuație de regresie este mai informativă decât un simplu coeficient de corelație. Spre exemplu, o cercetare asupra succesului în studiile de arhitectură stabilește

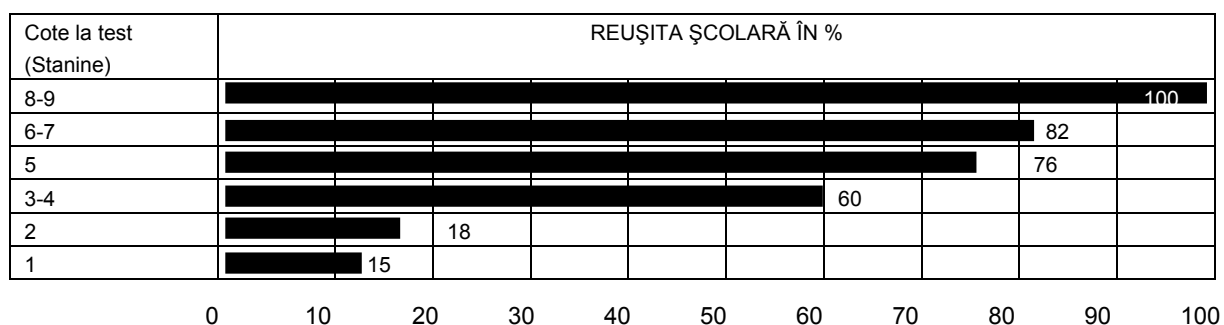


Figura 1.

---

ca fiind indicată o combinație între factorii intelectuali și de personalitate. Ecuația de regresie preconizată de un autor este următoarea:  $Y = 0,65 (Flx.) + 0,32 (G) + 0,17 (Ab. Arh.)$  (Strzalecki, 1973), în care *Flx.* indică flexibilitatea intelectuală, *G* denotă capacitatea de raționament iar *Ab. Arh.* înseamnă abilități în domeniul arhitecturii. Cu *Y* s-a notat predicția asupra succesului în obținerea diplomei.

2. Datele care servesc drept criteriu vor face obiectul unei descrieri complete și detaliate, arătându-se onest limitele lor, riscul de a fi afectate de factori parazitiți, sursele de eroare etc. Spre exemplu, când variabila-criteriu este rezultatul aprecierii se va arăta gradul de acord între evaluatori, precum și sursele de eroare identificate. În același timp se va specifica structura lotului de evaluatori (competența, formația lor, statutul profesional etc.).

3. Cotarea după criteriu va fi stabilită independent de rezultatele la teste, pentru a evita riscul de contaminare a criteriului. Persoanele care efectuează cele două operații trebuie să fie diferite; cu alte cuvinte, într-un studiu de validitate, persoana care aplică, notează și interpretează testul nu trebuie să știe despre subiecții examinați decât informațiile legate obișnuit de probă (altfel, apare efectul de halo). Tot așa notarea după criteriu, trebuie să satisfacă aceeași cerință, adică să aibă loc cu totul independent de testare.

4. Manualul va indica intervalul de timp ce s-a scurs între testare și momentul colectării datelor relative la criteriu. Dacă perioada respectivă este mai lungă se vor reține datele inițiale și cele finale. Se vor consemna, de asemenea, programele de instruire școlară sau profesională intercalate.

5. Se recomandă adunarea de date locale pentru stabilirea valorii predictive (a unei probe) decât să se utilizeze rezultate generale publicate. Se va insista asupra oportunității studiilor de validitate locală, propunându-se metodele necesare de lucru (tabele cu dublă intrare sau alte procedee).

6. Când persoana care utilizează testul are în vedere o modificare sensibilă în formatul, instrucțiunile, limbajul sau conținutul probei, atunci va trebui să valideze din nou instrumentul de măsură în noile condiții.

7. Este necesar să se insereze în manualul însoțitor toate datele cu privire la eșantionul de validare – mărimea sa, indici ai tendinței centrale și ai dispersiei – pentru ca

psihologul practician să-și dea seama dacă populația considerată de el prezintă caracteristici identice. Referitor la compoziția eșantionului se va indica, de asemenea, vârsta, sexul, statutul socio-economic, nivelul de instrucție, ca și orice alt factor, de diferențiere cu efecte sensibile (de exemplu, diagnosticul medical pentru loturi de bolnavi, experiența în producție pentru muncitori etc.).

8. Extrapolarea unor indici de la un grup la altul va avea loc în limitele unor caracteristici (de vârstă, nivel școlar sau profesional) identice sau foarte apropiate. Dacă în cadrul eșantionului studiat se conturează subgrupe cu indici de validitate diferiți, aceștia se vor stabili pentru fiecare subgrup separat sau se va face mențiunea că diferența nu există.

9. Un examen psihologic cuprinde de regulă, mai multe trăsături și aptitudini iar pentru evaluarea unei însușiri se aplică mai multe tehnici de studiere. Când informației-test i se adaugă alte date relevante despre subiect, care pot îmbunătăți prognoza, se va proceda la estimarea corelației multiple **R**, care să îmbrățișeze întreaga combinație de date accesibile. Dacă este cazul să se combine informațiile culese prin mai multe teste se va stabili de asemenea validitatea acestei combinații; când se propune un mod de ponderare (cote compozite) validitatea sa va fi verificată cu ajutorul unui alt eșantion decât cel inițial (contravalidare). O asemenea contravalidare este cu deosebire indicată atunci când numărul de predictorii cuprinși în cercetare este mai mare de 4-5 iar volumul eșantionului inițial sub 100.

10. Dacă vrem să întemeiem decizia pe o combinație complexă și neliniară de date se va verifica mai întâi superioritatea unei asemenea prognoze în comparație cu simpla combinație liniară. Utilizarea de variabile "moderatoare" este indicată cu deosebire în cazurile în care această tehnică mărește efectiv validitatea, determinată însă pe un eșantion de contravalidare. De asemenea, modificarea sistemului de notare sau de selecție a itemilor sugerat de un eșantion de încercare va fi validat pe un eșantion distinct sau lărgind grupul inițial de studiu. În general, orice soluție rezultată din simplul calcul pe hârtie trebuie validată experimental.

b) *Validitate de conținut.* Orice probă constituie un decupaj din activitatea subiectului, testul psihologic – spune Anastasi (1982) – reprezintă în esență o măsură obiectivă și standardizată a unui eșantion de comportament. Problema este ca acest decupaj sau

eșantion să fie reprezentativ; cu alte cuvinte, itemii ce compun testul să acopere – prin conținutul lor – ansamblul de sarcini, condiții sau procese mentale care privesc domeniul condensat prin test. Spre exemplu, când este vorba de probe de cunoștințe, de competență profesională sau probe de comportament social, problema eșantionării de conținut trece pe prim plan. Tot așa, în cazul inventarelor sau chestionarelor se pune întrebarea în ce măsură colecția de itemi acoperă ansamblul de situații sau aspecte pentru care chestionarul respectiv constituie un eșantion.

Este necesar de precizat – după Anastasi – că itemii unui test nu trebuie neapărat să se asemene strict comportamentului pe care testul îl explorează, este necesar numai ca o relație experimentală să fie dovedită între cele două aspecte.

Reprezentativitatea grupajului de itemi – în raport cu domeniul vizat de probă – nu poate fi stabilită prin inspecția sumară a conținutului. Este necesară o analiză logică de conținut – adeseori prin consultarea unui lot de specialiști – precum și date empirice – referitoare la proprietăți statistice ale materialului.

În acest sens se preconizează:

1. Manualul atașat testului trebuie să indice ansamblul de situații, comportamente sau procese psihice care sunt vizate prin test și în ce măsură se realizează o eșantionare adecvată (sau numai parțială) a acestora. De pildă, cu greu se poate numi “probă de imaginație” o activitate simplă de compunere de cuvinte din 3-4 litere date. Simplificarea în test nu trebuie să depășească o limită dincolo de care se pierd notele definitorii ale fenomenului examinat.

2. Când se consultă un lot de experți pentru a evalua în ce măsură itemii reuniți într-o probă constituie un eșantion satisfăcător dintr-un domeniu dat sau în ce măsură s-a făcut o cuantificare corectă a acestor itemi, atunci manualul va preciza experiența profesională și gradul de calificare al specialiștilor solicitați, precum și instrucțiunile (grila de categorii) pe baza cărora s-a lucrat.

3. Când itemii sunt aleși datorită lotului de specialiști este indicat să se arate gradul lor de acord sau de convergență, utilizându-se, spre exemplu, testul Kendall (când se cere ordinea itemilor) (I.Radu, 1993).

4. Pentru testele de achiziție școlară este necesar să se dea grila de clasificare adoptată pentru selecția itemilor. Se bucură de circulație în această privință tabele cu dublă intrare, în care se indică pe linii temele (conținutul

materiei verificate), iar pe coloane se înscriu obiectele, mai exact categoriile de sarcini sau operații mentale angajate în rezolvarea testului. Sunt utile în acest sens lucrările de taxonomie pedagogică sau modele sintetice ale inteligenței sau ale personalității.

5. Analiza de conținut, mai exact analiza de itemi servește și la stabilirea zonei discriminative a testului în raport cu subgrupele populației examinate, spre exemplu în raport cu vârsta subiecților. Se determină în felul acesta sensibilitatea probei, indiciu absolut necesar în utilizarea unui test.

6. Problema eșantionării itemilor sau a eșantionării conținutului se pune, de asemenea, în legătură cu alcătuirea formelor paralele ale unui test, când dispunem de o colecție mai mare de itemi și selectăm seturi paralele echivalente.

c) *Validitatea de construct* sau ipotetico-deductivă ne indică în ce măsură anumite concepte explicative sau anumite ipoteze permit înțelegerea și interpretarea rezultatelor obținute la o probă. Vorbim de teste de inteligență, de creativitate, de anxietate etc., postulăm deci o anumită teorie psihologică în spatele fiecărei probe. Examenul validității de construct comportă un demers logic și unul empiric; problema este aici de a testa sau controla teoria subiacentă probei pe baza unei acumulări progresive de date din surse variate. D. Groot (1962) observă în acest sens: trebuie luate în considerare toate rezultatele experimentale care se referă la contextul teoretic căruia îi aparține conceptul, la rețeaua nomologică învecinată. De exemplu, când vorbim de “caracterul neurotic” sau de “neurotizism” vom considera întregul ansamblu de date și ipoteze referitoare la temă. Practicianul s-ar putea eventual mulțumi cu un indice semnificativ de validitate predictivă care răspunde obiectivului său nemijlocit. Nu putem admite însă o testare oarbă, ruptă de achizițiile științei psihologice, care trebuiesc încorporate în construcția probelor. Orice test măsoară/explorează ceva care este definit printr-un concept (*construct*), orice instrument de diagnostic realizează în fapt o definiție operațională, propune adică o noțiune implicită asupra unui fenomen psihic (aptitudine, însușire etc.) și problema se pune dacă o asemenea noțiune însumează datele cercetării psihologice sau constituie un simplu clișeu preluat din conștiința comună. Toți indicii de validitate tind să devină contribuții la evaluarea validității de construct. O probă este predictivă întrucât cuprinde adevărul științific,

reușita practică a unui instrument de psiho-diagnostic este dovada justetei sale. Tot așa, conținutul probei este o ilustrare a conceptului. Cele trei tipuri de validitate nu sunt independente. Studiul exhaustiv al unei probe trebuie să înglobeze în mod normal toate tipurile de validitate.

Pentru controlul empiric al validității de construct a unei probe se recomandă corelarea cu alte teste deja verificate pentru un anumit domeniu (validitate concurrentă).

De notat că probele nu sunt psihologic pure. În acest sens se recurge la analiza factorială pentru a preciza saturația în factorul presupus a sta la baza alcătuirii probei. În documentul însoțitor se va indica partea din varianța rezultatelor la test datorită factorului de referință (vizat prin proba respectivă).

În sfârșit, proceduri experimentale bazate pe eșantioane contrastante (de exemplu, manageri de succes – manageri cu eșec etc.), pot fi în măsură să conducă la concluzii certe în problema validității conceptuale.

★

A două calitate psihometrică a probei o constituie *fidelitatea*.

Termenul de fidelitate desemnează în genere acordul sau stabilitatea așteptată a măsurărilor analoge. Calculul fidelității - în fapt un calcul de corelație – presupune în mod curent compararea a cel puțin a două colecții de măsurări obținute prin:

- "forme paralele" ale aceleiași probe,
- repetarea testului după un timp,
- aplicarea unei singure probe (corelații între rezultate parțiale, consistența inter-itemi etc.).

Teoretic, "cota adevărată" a unui subiect rămâne constantă de la o aplicare a testului la alta, de la o formă sau parte a testului la alta; practic însă, "cota observată" variază într-o anumită măsură. Această fluctuație sau variație a notei unui subiect în seria de aplicare sau în cadrul unei singure aplicări se numește eroare. Evident, sursele de eroare sunt mai multe: variații în eficiența fiziologică a individului, în nivelul de motivare, în conținutul probei și a condițiilor extreme de testare etc. Un studiu asupra fidelității probei estimează componentele varianței cuprinse sub numele de eroare (varianță reziduală), tehnicile de determinare fiind cunoscute.

Ca cerințe generale se preconizează:

1. Materialul de îndrumare ce însoțește o probă va insera în chip obligator datele necesare asupra fidelității, precizând tehnica pe baza căreia s-au obținut: test-retest, par-impar, forme paralele, corelații inter-itemi sau între rezultate parțiale, precum și eșantionul. Se vor indica valorile ca atare ale coeficienților de fidelitate (nu pragul de semnificație  $p$ ), de asemenea se va arăta eroarea standard a unei măsurări individuale ( $E_{m\grave{a}s}$ ) ( G. M. Smith, 1971). Prezentarea datelor se va face pe cât este posibil, pe componente ale varianței reziduale: fluctuații sau contradicții în răspunsurile subiectului, eterogenitatea itemilor (situațiilor) variații în administrarea probei, deseori între corectori etc. Cel puțin unul din aspectele fidelității este necesar să fie prezentat.

2. Cerințele cu privire la fidelitate privesc nu numai indicatori de ansamblu, ci și indici parțiali sau orice combinație de rezultate (sumă, diferență, cât) pe care o probă îi recomandă într-un scop sau altul.

3. Cu privire la eșantionul de studiere a fidelității trebuie menționați toți factorii selectivi aflați în conexiune cu variabila măsurată; când testul este destinat clasificării sau discriminării în funcție de anumite categorii se vor calcula indici de fidelitate separat pe fiecare subgrup. De exemplu, un subtest de "raționament mecanic" dintr-un test de aptitudini nu prezintă aceeași fidelitate pentru băieți sau fete; în consecință se vor face diferențierile de rigoare.

4. Indicii de fidelitate nu probează automat validitatea testului, deși rezultatele care nu satisfac exigențele fidelității nu pot fi valide.

5. Când avem de-a face cu două forme ale unei probe, aplicabile aceleiași populații, echivalența lor nu va fi admisă decât cunoscând mediile, abaterile standard și coeficientul de corelație pentru eșantioanele studiate. Se va menționa de asemenea și intervalul de timp între testările practicate.

6. Când proba vizează o trăsătură omogenă, va trebui să se verifice consistența internă a probei, făcând apel la metodele "prin înjumătățire" sau formula Kuder-Richardson (Vezi Anexa B).

7. În ceea ce privește compararea în timp a rezultatelor, stabilitatea lor, manualul atașat testului va trebui să ofere informații corespunzătoare, ținând seama de faptul că intervalul de timp între testări succesive capătă semnificație în funcție de natura probei și de interpretarea propusă. În genere, o probă ale

cărei rezultate variază de la o zi la alta este pusă sub semnul întrebării.

★

În sfârșit, o seamă de cerințe este bine să fie amintite cu privire la administrarea și etalonarea probelor.

1. Psihologul practicant va trebui să urmeze fidel instrucțiunile care au servit la stabilirea normelor și a datelor relative la validitate. Orice modificare în modul de administrare a probei se răsfrânge asupra rezultatelor, inclusiv asupra etalonului. Instrucțiunile vor trebui să fie redactate fără echivoc astfel încât subiecții să înțeleagă sarcina de îndeplinit în perspectiva prevăzută de autor. Spre exemplu, în inventare de personalitate se cere subiectului să dea primul răspuns care îi vine în minte; de asemenea, în chestionare de interese se vor preciza condițiile de opțiune a unei preocupări: pentru o situație ocazională sau sistematică, pentru o situație ideală sau una reală, marcată de constrângerile externe.

2. Se va prevedea un mod de corelație pentru șansa de a se obține răspunsuri date la întâmplare ("pe ghicite") la itemii cu alegere multiplă sau cu răspunsuri dihotomice.

3. În întocmirea tabelului de norme se vor utiliza în general cote standardizate și anume: decile, centile sau cote normalizate (I. Radu, 1993). De largă răspândire sunt așa numitele cote **T** având  $m = 50$  și  $\sigma = 10$  pentru cote standardizate de două cifre, iar pentru cote de o singură cifră (G. M. Smith, 1971) sunt indicate cote în genul staninelor având  $m = 5$  și  $\sigma = 2$ .

4. În practica testării, etalonările locale sunt adesea mai importante și mai utile decât cele publicate, motiv pentru care se va insista în manualul atașat testului asupra tehnicilor recomandate de etalonare.

5. Tabelele de norme vor indica volumul eșantionului de etalonare, precum și indici adecvați asupra tendinței centrale (media sau mediana) și dispersiei (abaterea standard sau abaterea semiintercvartilă). Vor fi înserate, de asemenea, toate datele asupra compoziției eșantionului, precizându-se factorii de selecție (vârstă, sex, formație școlară etc.) și dacă aceștia influențează rezultatele (în ce măsură).

6. Se recomandă ca textul însoțitor să arate clar condițiile etalonării: modul de testare, nivelul motivării, atitudinii față de examinare ș.a. Una este să etalonezi, spre exemplu o probă pe candidații la angajare, la

examen de admitere – deci o situație de selecție – și alta este condiția de consiliere profesională.

★

În partea de încheiere, trebuie subliniată din nou necesitatea și importanța unei formații psihologice avansate în întocmirea și realizarea oricărui program de testare. Este vorba de stăpânirea unui minim de cunoștințe despre principiile testării, despre erorile de măsurare, despre modul de interpretare a datelor etc. știind că același rezultat comportă explicații alternative, răspunzând aceluiași exigențe etc. Este de așteptat ca psihologul calificat să fie înarmat cu criteriile sau normele necesare care să-l ajută să discearnă mijloacele ce îi stau la dispoziție, să aleagă din repertoriul clasic probe adecvate în raport cu situația completă, să interpreteze corect datele obținute, având conștiința lucidă, critică asupra valorii și limitelor fiecărui instrument de lucru.

## ANEXE

### A. COEFICIENT DE CORELAȚIE BISERIAL ( $r$ bis)

În studiile referitoare la validitatea unui test se întâlnesc situații în care variabila-criteriu se prezintă numai sub două variante sau valori: admis/respins, bun/slab etc.

#### *Exemplu*

Înainte de un examen de admitere la un curs de instruire profesională candidații au fost supuși unei testări psihologice. Cotele, obținute la aceste probe, merg de la 0 la 10 puncte. Pe de altă parte, la încheierea examenului de admitere rezultatele finale apar în formă dihotomică: o parte din candidați au fost declarați admiși iar ceilalți respinși. Notăm cotele la test cu  $x$  iar variabila criteriu cu  $y$ . se observă ușor că variabila-test ia valori între 0 și 10 în timp ce variabila-criteriu este dihotomizată: admis/respins. Datele obținute sunt sintetizate în tabelul 1 (după J.M. Faverge).

Se poate stabili proporția celor admiși în urma examenului:

$$p = \frac{59}{113} = 0,52$$

De asemenea proporția celor respinși:

$$q = \frac{54}{113} = 0,48$$

Se observă imediat cum  $q = 1 - p$ .

Din același tabel se poate urmări ce rezultate au obținut în cazul testării psihologice candidații declarați admiși, se face o medie a acestor rezultate, notată cu  $m'$ . De asemenea, se poate face același lucru pentru categoria celor respinși, notând media la test cu  $m''$ .

Formula de calcul este următoarea:

$$r_{\text{bis}} = \frac{m' - m''}{\sigma} \cdot \sqrt{pq}$$

Notațiile sunt deja precizate; se menționează în plus că  $\sigma$  este abaterea standard a rezultatelor (cotelor) obținute în cadrul testării psihologice a întregului lot de candidați.

În exemplu dat:

$$\text{media celor admiși } m' = \frac{359}{59} = 6,08$$

$$\text{media celor respinși } m'' = \frac{241}{54} = 4,46$$

abaterea standard pe ansamblu  $\sigma = 1,88$

Înlocuit în formulă vom avea:

$$r = \frac{6,08 - 4,46}{1,88} \cdot \sqrt{0,52 \times 0,48} = 0,86 \times 0,5 = 0,43$$

Când într-un studiu de validare, criteriu ( $y$ ) ne apare sub forma unei clasificări trihotomice – de exemplu, buni-mijlocii și slabi – în timp ce datele testării psihologice ( $x$ ) se prezintă într-o formă mai nuanțată – ca în exemplu precedent – atunci se utilizează pentru calculul corelației o formulă analoagă propusă de C. Burt:

$$r_{\text{tris}} = \frac{m' - m''}{\sigma} \cdot \frac{1}{\frac{z'}{p'} + \frac{z''}{p''}}$$

în care:

$m'$  și  $m''$  reprezintă mediile la test ( $x$ ) pentru cei buni și cei slabi;

$p'$  și  $p''$  proporția de buni, respectiv de slabi ( $p' + p'' < 1$ );

$z'$  și  $z''$  sunt ordonatele curbei normale reduse potrivit împărțirii între categorii;  $\sigma$  este abaterea standard a distribuției complete a datelor la test ( $x$ ).

Valorile rapoartelor  $\frac{z}{p}$  sunt date în anumite

tabele speciale. Figura nr. 2 ilustrează sensul notațiilor utilizate. Se observă că grupul celor mijlocii nu intră în calcule.

Tabelul nr. 1

Cote la test \ Rezultate la examen	0	1	2	3	4	5	6	7	8	9	10	Total
Admiși	0	1	0	3	5	10	16	14	6	3	1	59
Respinși	1	2	4	7	13	12	9	4	2	-	-	54
Total	1	3	4	10	18	22	25	18	8	3	1	113

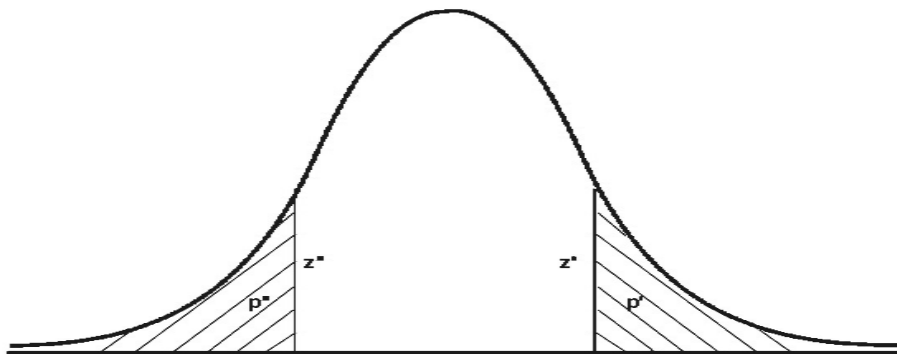


Fig. 2

## B. Formula 20 KUDER-RICHARDSON (după A. Anastasi)

Este o tehnică de studiere a configurației răspunsurilor date de subiecți la toți itemii unui test, pornind de la datele obținute în cadrul unei singure administrări. Este vorba de itemi care se cotează doar prin 1 sau 0, iar proba ca atare nu comportă timp limitat. Gradul de consistență interitemi se determină pe baza examinării rezultatelor la fiecare item, utilizându-se formula:

$$r_{1I} = \left( \frac{n}{n-1} \right) \frac{\sigma_t^2 - \Sigma_{pq}}{\sigma_t^2}$$

în care:

$r_{1I}$  este coeficientul de fidelitate pentru întregul test;

$n$  reprezintă numărul de itemi reuniți în test;

$\sigma_t^2$  este dispersia calculată pe ansamblul cotelor totale la test.

Termenul nou în formulă este  $\Sigma_{pq}$ . Aceasta se află prin întabelarea proporției de persoane care au reușit ( $p$ ), în contrast cu cei care au întrunit cota 0. Se calculează produsul  $pq$  pentru fiecare item în parte și apoi se totalizează aceste produse pentru toți itemii, obținându-se  $\Sigma_{pq}$ . Indicele obținut ne arată gradul de omogenitate a itemilor și totodată măsura în care s-a realizat o sensul congruenței sau consonanței în răspunsurile subiecților.

## Bibliografie

- Anastasi, A. (1982). *Psychological testing*. Londra: Mac Millan Comp.
- Groot, D. (1962). *Quelques reflexions sur les problemes de la psychometrie. Les problemes de la mesure en psychologie*. Paris: P. U. F.
- Kulcsar, T. (1978). *Factorii psihologici ai reușitei școlare*. București: Editura Didactică și Pedagogică.
- Radu, I. (1974). *Psihologie școlară*. București: Editura științifică.
- Smith, G.M. (1971). *Ghid simplificat pentru psihologie și pedagogie*. București: Editura Didactică și Pedagogică.
- Standards for Educational and Psychological Tests*, American Psychological Association, Inc. , 1974.
- Strzalecki, A. (1973). *Prediction of academic succes in architectural school as in preparatory study*

*to become designer*. Warsaw: Metodologia projektowania inzynierskiego.

Szomoskozi, Șt. (1997). *Evaluarea potențialului intelectual la elevi*. Cluj-Napoca: Presa Universitară Clujeană.