

METODOLOGIA EVALUĂRII CALITĂȚII VIETII ÎN MEDIUL EDUCAȚIONAL: ALGORITMUL DE ELABORARE A INSTRUMENTELOR DE EVALUARE

Carolina PLATON, Zinaida BOLEA, Tatiana TURCHINĂ

Catedra Psihologie Aplicată

The article reflects the author's opinions towards the algorithm of elaboration of assessment tools. The concrete steps are proposed to undertake in order to elaborate, to adjust and to validate the assessment tools: determination of functions, defining of concepts, elaboration of scales, evaluation of scales and validation of tools.

Orice domeniu științific aspiră să măsoare și să descrie într-un mod cât mai exact fenomenele pe care le studiază, în special când se recurge la instrumente pentru a evalua o achiziție sau abilitate, o calitate sau caracteristică, o opinie sau atitudine. În contextul celor expuse, putem evidenția o dublă incursiune în metodologia elaborării instrumentelor de evaluare, a căror cunoaștere este indispensabilă procesului de elaborare a oricărui instrument. Prima se referă la *problema algoritmului de elaborare a instrumentelor de evaluare*. A doua ține de *problema expertizei și a criteriilor de selecție a experților*, rolul cărora în elaborarea și validarea instrumentelor de evaluare este de necontestat.

Reperetele teoretice expuse în articolul de față reflectă prima incursiune: *problema algoritmului de elaborare a instrumentelor de evaluare*. Precizăm că în intenția noastră nu a fost să prezentăm un concept general al construirii instrumentelor și nici să detaliem această temă. Scopul explicit este de a evidenția reperetele teoretice pe care ne-am bazat în cursul elaborării/adaptării/validării instrumentelor de investigare a dimensiunilor psihologice ale calității vieții în mediul educațional.

Una dintre problemele esențiale ale construirii instrumentelor de evaluare este determinarea algoritmului acestui proces. Literatura de specialitate ne prezintă mai multe exemple de „algoritmizare”, care nu diferă esențial de la un autor la altul.

În prezenta cercetare ne-am condus de algoritmul autorilor Laveault și Gregoire [1], care disting cinci etape în elaborarea instrumentelor de evaluare:

- 1) determinarea funcțiilor prevăzute de instrumentul de evaluare;
- 2) operaționalizarea conceptului;
- 3) elaborarea itemilor;
- 4) evaluarea itemilor;
- 5) validarea instrumentului.

1) Determinarea funcțiilor instrumentelor de evaluare

Prima etapă s-a referit la determinarea funcțiilor pe care instrumentul trebuie să-l îndeplinească, răspunzând la întrebarea *la ce va servi?* În contextul prezentei cercetări, mai multe distincții funcționale ale instrumentelor au fost trasate.

O distincție netă se referă la instrumentele de evaluare a componentelor structurale ale calității vieții: bunăstarea fizică și materială, bunăstarea emoțională, afirmarea personală și integrarea socială. Instrumentele au fost construite/selectate/adaptate în mod diferit în dependență de funcția pentru care sunt destinate.

În primul caz ne-am condus, în temei, de metodologia „obiectivă”, care se bazează pe date colectate indirect (asistența medicală, locuința, veniturile, consumul).

În celelalte cazuri ne-am condus, de regulă, de metodologia „subiectivă”, care se bazează pe date colectate direct de la persoanele evaluate (satisfacerea nevoilor-cheie, compatibilitate intelectuală, similaritate cognitivă, potrivire temperamentală, competență profesională, echitate și responsabilitate interpersonală etc.).

Distincția stabilită între metodologia „obiectivă” și „subiectivă” nu este singura. În domeniul evaluării există profunde diferențe între instrumentele normative și cele criteriale. Instrumentele de evaluare normativă vizează discriminarea subiecților unei populații pentru care este construit instrumentul. Ele pot furniza informații privind gradul de manifestare a unei caracteristici a subiectului în raport cu nivelul acestor caracteristici prezente populației în ansamblu. De exemplu, care este nivelul de „manifestare/percepere” a calității vieții în diverse regiuni ale Moldovei în raport cu nivelul mediu prezent la ansamblul populației? Din contra, instrumentele criteriale au drept funcție să determine dacă un subiect posedă sau nu caracteristici luate drept refe-

rință. În acest caz, subiecții, prin randamentul obținut, se împart în cei care depășesc/nu depășesc acest criteriu. De exemplu, care este nivelul de „manifestare/percepere” a calității vieții în raport cu un criteriu de referință (Uniunea Europeană, România etc.)?

Așadar, funcțiile prevăzute pentru un instrument de evaluare determină esențial caracteristicile lui. Totodată, considerăm puțin reală crearea unor instrumente „generalizatoare” care ar putea să răspundă concomitent la toate necesitățile practicii evaluative.

2) Operaționalizarea conceptului

Operaționalizarea conceptului presupune definirea cu precizie a caracteristicilor pe care instrumentul vrea să le măsoare. Pentru a trece de la o intenție mai vagă la definirea operațională a conceptului au fost utilizate două metode.

- Analiza de conținut a convorbirilor. Metoda presupune interviuarea persoanelor care aparțin populației vizate de instrument. Un interviu liber sau semistrukturat permite colectarea numeroaselor informații care, ulterior, se clasifică prin intermediul analizei de conținut.
- Referirea la un model teoretic. Spre deosebire de metodele care pornesc de la experiență, aceasta se bazează pe modelul teoretic fundamentat în cursul cercetării.

Or, în elaborarea instrumentelor utilizate în prezenta cercetare ne-am bazat atât pe teoriile existente, cât și pe multiplele date empirice din vastul domeniu al evaluării.

3) Elaborarea itemilor

Cu circa 60 de ani în urmă, Gallup, fondatorul celebrului Institut de Sondaje, afirma că, spre deosebire de atenția sporită acordată constituirii eșantionului, atenția acordată elaborării itemilor a rămas destul de modestă. Totodată, diferențele în construirea itemilor conduc la rezultate cu variații mai mari decât cele observate în funcție de diferite tehnici de eșantionare. Această constatare a autorului își păstrează actualitatea și poate fi raportată la itemii construiți pentru instrumentele utilizate în științele psihologice [2].

Lucrările privind elaborarea itemilor au cunoscut o eclipsă destul de lungă, în timpul căreia cercetările au fost axate mai mult pe caracteristicile subiecților. Metodologia utilizată în aceste lucrări este, în temei, empirică: se variază sistematic caracteristicile itemilor propuși diferitelor grupe de subiecți și se examinează diferențele între răspunsuri. Aceste studii experimentale se completează cu analize suplimentare ale datelor provenind din anchete largi de opinii, prin care se încearcă să se pună în evidență caracteristicile itemilor care au influență asupra calităților metrice.

În opinia savanților [3], la etapa actuală este greu de propus un ansamblu exhaustiv de reguli de elaborare a itemilor. Ce format de itemi să alegem? Câți itemi trebuie elaborați? Care trebuie să fie gradul de dificultate a itemilor? Iată doar câteva întrebări care solicită răspuns.

Astfel, dat fiind faptul că elaborarea itemilor ridică numeroase probleme, devine indispensabilă respectarea rigorilor metodologice, acestea referindu-se, cu precădere, la *A)* dispozitivele de prezentare și *B)* dispozitivele de răspuns.

A) Influența dispozitivelor de prezentare vizează, de regulă, chestionarele (atitudini, reprezentări, opinii etc.), interesul cercetătorilor focalizându-se asupra unui șir de efecte: a) semantice, b) formale, c) contextuale, d) numerice.

a) Studiile privind efectele semantice arată că schimbări foarte lejere în alegerea termenilor utilizați pot modifica sensul itemului și, în consecință, răspunsurile subiecților. Exemplul cel mai cunoscut se referă la asimetria interdicție-autorizare [4]. De exemplu, când americanii sunt întrebați dacă Statele Unite trebuie să interzică discursurile contra democrației, 21% răspund afirmativ și 79% răspund negativ. Când ei sunt întrebați dacă Statele Unite trebuie să autorizeze discursurile contra democrației, 52% răspund afirmativ și 48% răspund negativ. Or, înțelegerea proceselor legate de răspunsurile la chestionare justifică utilizarea unui anume termen în elaborarea itemilor.

b) Efectele formale ale itemilor ocupă un loc mai modest în literatura de specialitate. Analiza studiilor la acest capitol [5] ne-a permis să desprindem un șir de constatări. Astfel, caracteristicile formale au o influență majoră asupra răspunsurilor subiecților și a calităților metrice ale itemilor. Iată de ce itemii formulați negativ care induc tendința de a răspunde mai puțin distinct, micșorând, astfel, fidelitatea scalei, au fost evitați; redactarea itemilor a fost cât mai explicită; utilizarea itemilor socialmenete dezirabili/nedezirabili și referințele la termeni precum „frecvent”, „de obicei”, „adesea” etc. au fost maximal evitate.

c) Efectele contextuale figurează, de asemenea, într-un șir de lucrări de specialitate [6]. Ele se referă la instrument în ansamblu și arată că prezența unui item poate modifica răspunsul pe care subiectul l-ar fi dat la un alt item dacă primul nu ar fi expus. Din această perspectivă, suscită interes locul unui item în raport cu altul, aranjamentul lor în chestionare.

d) Efectele numerice ale unui instrument depind de mai mulți factori. Primul factor se referă la economicitatea aplicării instrumentului. În dependență de intenția de a construi un instrument scurt care ar putea fi aplicat în 10 minute sau un instrument mai amplu care ar necesita mai mult timp, numărul de itemi variază considerabil, însă nu dincolo de limita unor valori minimale care atribuie siguranță pentru criteriile de apreciere a instrumentului. În literatura de specialitate doar Kline, citat de Dickes [7], face referință la numărul minim de zece itemi pe care trebuie să-l conțină un test, număr considerat mic de alți autori [8]. Al doilea factor vizează nivelul dorit de fidelitate a instrumentului. Un instrument cu mulți itemi este mai fidel decât un instrument cu puțini itemi. Al treilea factor ține de eliminarea aproape inevitabilă a anumitor itemi după evaluarea de către experți și pretest. Dacă se vrea ca versiunea finală să conțină un anumit număr de itemi, pentru versiunea inițială se selectează un număr mai mare de itemi decât cel dorit. De exemplu, pentru varianta finală de 15 itemi noi am selectat 25-30, dintre care am reținut 15 itemi mai reprezentativi. De regulă, un surplus de 30-50% de itemi este necesar pentru a evita surprizele neplăcute care pot interveni după pretest.

B) *Influența dispozitivelor de răspuns* se manifestă prin formatul itemilor. Se disting formate pentru chestionare și formate pentru testele de performanță. Dat fiind faptul că ambele sunt utilizate în prezenta cercetare, specificăm caracteristicile lor.

Evaluarea atitudinilor, opiniilor, reprezentărilor, valorilor, intereselor, trăsăturilor de personalitate face apel la formate specifice chestionarelor. Trei formate sunt, de regulă, utilizate: a) itemi dihotomici, b) itemi categoriali bipolari, c) itemi cu alegere forțată.

a) Itemii dihotomici sunt constituiți dintr-o propoziție în raport cu care subiectul trebuie să exprime acordul sau dezacordul. Alegerea poate fi făcută între „da – nu” „adevărat – fals”, „acord – dezacord”. Construirea itemilor dihotomici este, în aparență, simplă. Totuși, acest format ridică un șir de probleme. O problemă ține de formularea enunțului. De exemplu, o formulare de tipul „Cunoscții mei nu mă înțeleg” poate condiționa două răspunsuri: „Da, cunoscții mei nu mă înțeleg” sau „Nu, cunoscții mei nu mă înțeleg”. A doua versiune corespunde unei formulări mai firești decât prima; totuși, subiectul care este de acord cu enunțul trebuie să aleagă „Da”. Utilizarea modalităților de răspuns „adevărat-fals” sau „acord-dezacord” poate reduce într-o anumită măsură această problemă. O altă problemă vizează caracterul marcat al itemului. De exemplu, în raport cu enunțul „Profesorul este o persoană relativ obiectivă”, interpretarea termenului „relativ” poate varia și poate antrena, în dependență de această interpretare, răspunsuri diferite. Încă o problemă ridicată de itemii dihotomici ține de fenomenul dezirabilității sociale. Numeroși subiecți au tendința să mascheze veritabilă lor alegere și, din contra, să selecteze alegerea opusă, dat fiind faptul că aceasta este socialmente mai valoroasă. Tendința evidențiată poate să decurgă din dorința subiectului de a apărea într-o lumină favorabilă vizavi de opinia pe care evaluatorul și-o poate crea despre el. Pentru a evita momentele datorate dezirabilității sociale, itemii trebuie să fie indicatori valizi, fiind în același timp socialmente acceptabili.

b) Itemii categoriali bipolari intervin în anumite condiții de evaluări, când e nevoie de a da răspunsuri mai nuanțate decât cele dihotomice. Ansamblul alegerilor, în asemenea cazuri, constituie categorii ordonate sau scale care variază în dependență de mai multe momente: numărul de trepte, etichetarea treptelor, prezența unui răspuns median etc.

Una dintre întrebările „de serviciu” privind utilizarea scalei de răspuns se referă la numărul de trepte. Conform studiilor [9], numărul optim se situează între cinci și nouă trepte. Asemenea scale au o bună fidelitate, limitează subiecții indeciși și discriminează capacitățile lor.

Etichetarea treptelor este deseori de natură numerică, însă alegerea ei nu este neutră. Astfel, după cum demonstrează Schwager [10], rezultatele diferă în dependență de faptul ce scală a fost propusă subiecților: „0 – 10” sau „-5 – +5”. În primul caz, 34% dintre subiecți au ales un răspuns între 0 și 5, în timp ce în al doilea caz numai 13% au ales un răspuns între -5 și 0. Pare că subiecții interpretează treapta de jos în scala „0 – 10” doar ca o absență a reușitei, pe când în scala „-5 – +5” această treaptă este interpretată ca prezența unor importante eșecuri.

Pentru a evita problemele generate de etichetele numerice, acestea pot fi înlocuite cu etichete verbale. Însă, și această soluție ridică un șir de probleme. Un lucru pare a fi stabilit: etichetele pot schimba semnificația întrebării și epuizează problema interpretării scalei de către subiecți. Savanții Pace și Friedlander (citați după Cronbach [11]) arată că printre persoanele care au răspuns „deseori” pe o scală de frecvență etichetată „niciodată”, „ocasional” „deseori” și „foarte des”, 41% au efectuat activitatea în cauză o dată pe săptămână, 33% – de una-două ori pe săptămână, iar 26% – o dată în lună.

Și prezența unui răspuns median într-o scală (de exemplu: „acord total”, „acord parțial”, „neutru”, „dezacord parțial”, „dezacord total”) ridică un șir de întrebări. Existența unei asemenea posibilități are un anumit avantaj: ea permite subiecților indeciși să aleagă un răspuns în loc să se abțină de la el. Problema pusă de prezența sau

absența unei modalități de răspuns de tipul „nu știu”, „fără opinie” este, practic, de același ordin ca și precedenta. Dacă subiectului i se oferă, în mod explicit, oportunitatea să nu răspundă, pe de o parte, se poate mări numărul abținerilor de la răspuns, însă, pe de altă parte, includerea acestor modalități este o precauție indispensabilă pentru evitarea răspunsurilor întâmplătoare. Molenaar [12] vorbește în asemenea cazuri despre o „presiune spre a răspunde” și arată că subiecții care nu au oportunitatea de a se abține de la răspuns dau răspunsuri totalmente fictive.

c) Itemii cu alegere forțată au fost creați pentru a încerca să se soluționeze una dintre problemele evocate anterior: influența dezirabilității sociale asupra alegerii răspunsurilor. Un item cu alegere forțată este constituit, de regulă, din patru elemente: doi indicatori valizi și doi indicatori nonvalizi ai unei caracteristici. Unul dintre indicatorii valizi este socialmente dezirabil, altul – socialmente nedezirabil. Același lucru este caracteristic și pentru itemii nonvalizi. Din cele patru elemente, subiectul alege caracteristica care corespunde cel mai mult și cel mai puțin opiniei sale. Astfel, impactul dezirabilității sociale se reduce considerabil.

Formatele de itemi pentru testele de performanță diferă de la un autor la altul. Tipologia clasică distinge două mari tipuri de formate: răspunsuri închise (itemi cu alegere multiplă, itemi dihotomici, itemi de cuplare), în care subiectul trebuie să aleagă printre modalitățile de răspuns impuse, și răspunsuri deschise (itemi cu răspuns scurt, itemi cu răspuns narativ constrâns, itemi cu răspuns narativ desfășurat), în care subiectul este liber să elaboreze răspunsul său.

În cadrul răspunsurilor închise pot fi făcute distincții mai fine, în funcție de tipul de alegere cerut subiecților. Itemii cu alegere multiplă solicită o alegere din mai multe opțiuni, care, în afară de soluția corectă, cuprinde și soluții incorecte. Numărul alternativelor poate varia, însă, în majoritatea cazurilor, el este cuprins între patru și șase. Itemii dihotomici (adevărat – fals, da – nu, acord – dezacord) comportă un singur enunț, a cărui veridicitate trebuie evaluată de către subiect. Itemii de cuplare sunt o formă derivată din enunțurile cu alegere multiplă. Subiectului i se cere să pună în relații un număr de enunțuri cu un număr respectiv de răspunsuri corecte.

Deși incluse în răspunsurile deschise, itemii cu răspuns scurt posedă, deseori, caracteristicile itemilor închiși. Numiți și itemi „obiectivi”, ei cer subiecților să furnizeze un cuvânt sau o frază care poate fi evaluată ca corectă sau incorectă, fără ca subiectivitatea evaluatorului să intervină. Însă, nu orice item cu răspuns scurt este și obiectiv. Unii itemi cer din partea evaluatorului o judecată privind calitatea răspunsului, fapt ce diminuează fiabilitatea lor. Problema subiectivității evaluatorului apare cu mult mai pregnant în itemul cu răspuns narativ, care poate fi constrâns sau desfășurat. În primul caz, itemul conține anumite directive. În al doilea caz, subiectului i se lasă toată libertatea în structurarea răspunsului.

Astfel, obiecțiile privind formatele de răspunsuri închise/deschise țin de două momente esențiale: itemii închiși constrâng subiectul să adopte cadrul de referință ales de evaluator; itemii deschiși cauzează complexitate și subiectivitate în cotare. Totuși, grație simplității, itemii închiși sunt mai utilizați atât în probele de performanță, cât și în diverse tipuri de chestionare.

4) Evaluarea itemilor

O determinare exactă a funcțiilor și o metodologie riguroasă a elaborării itemilor sunt condiții necesare, dar nu și suficiente, pentru a obține un instrument valid și fiabil. Pentru a asigura proprietățile metrice ale itemilor, o evaluare minuțioasă a acestora trebuie să fie efectuată. Două demersuri complementare au fost urmate pentru a realiza această sarcină: a) evaluarea itemilor de către experți și b) pretestarea itemilor.

a) Evaluarea itemilor de către experți presupune o validare de conținut care constă în a aprecia în ce măsură itemii instrumentului sunt reprezentativi pentru conceptul vizat. Orice demers riguros de validare trebuie să dezbuteze cu o definiție exactă a conceptului. În mod generic, termenul „concept” desemnează ceea ce este vizat de instrument. Totodată, un instrument nu măsoară direct o realitate, ci, mai degrabă, o reprezentare sau un model al acesteia. De exemplu, un test de inteligență permite doar evaluarea modelului inteligenței promovată de autor. Același remarcă se referă la instrumentele destinate să măsoare alte „realități”, de exemplu, calitatea vieții etc. În toate cazurile nu realitatea absolută, ci o anumită concepție, reprezentare a realității este evaluată.

Așadar, calitatea validării de conținut depinde mult de precizia cu care a fost definit conceptul și de acordul experților vizavi de fațetele acestui concept. Mărirea „fațetă” poate desemna competențe, categorii comportamentale, capacități și/sau trăsături de personalitate etc. Definirea clară a conceptului permite identificarea fațetelor care trebuie luate în considerare pentru a selecta itemii instrumentului. Ea ne permite, de asemenea, să specificăm variabilele care nu fac parte din concept. Un concept definit vag nu permite obținerea unei validități satisfăcătoare a instrumentului elaborat.

Indicii de validitate a instrumentului sunt convenționali, dat fiind faptul că ei depind de definirea conceptului. Caracterul convențional al validității de conținut ne conduce la ideea că validitatea nu este absolută. Validarea

de conținut a unui instrument rămâne întotdeauna relativă în raport cu timpul și locul unde a fost realizată. Drept urmare, ea trebuie periodic reevaluată. Validarea de conținut a instrumentelor elaborate în cadrul prezentei cercetări a fost realizată de un grup de experți în conformitate cu principiile de bază prezentate în Tabel [13].

Tabel

Principiile de bază pentru validarea de conținut

1. A defini domeniul și fațetele (dimensiunile, factorii) conceptului.
2. A utiliza experți și reprezentanți ai populației de referință pentru a elabora itemii instrumentului.
3. A supune toate aspectele instrumentului validării de conținut.
4. A utiliza mai mulți experți și a cuantifica judecățile lor pentru a valida conținutul.
5. A examina reprezentarea proporțională a itemilor în raport cu diferite fațete ale conceptului.
6. A prezenta rezultatele validării de conținut.
7. A lua în considerare toate analizele ulterioare pentru a perfecționa validarea de conținut.

Din rele relatate reiese că judecățile experților joacă un rol esențial în procedura validării de conținut a unui instrument de evaluare.

b) Pretestarea itemilor vine să completeze evaluarea itemilor de către experți. Pretestarea permite colectarea nemijlocită a datelor empirice în cadrul populației de referință și constă în prezentarea itemilor unui eșantion al acestei populații. Mărimea eșantionului depinde de gradul de omogenitate a populației vizate de test. Dacă populația este omogenă, un eșantion de circa 50 de subiecți permite o evaluare satisfăcătoare a itemilor [14]. Rezultatele sunt analizate din punct de vedere calitativ și cantitativ. Comentariile subiecților cu privire la itemi pot contribui la reformularea și perfecționarea enunțurilor. Concomitent, în cadrul pretestării pot fi soluționate și alte probleme: de administrare a instrumentului, de înregistrare a răspunsurilor, de cotare a rezultatelor etc. Fiind ignorate, problemele evocate pot conduce la diminuarea considerabilă a validității instrumentului. În plus, pretestarea itemilor permite realizarea analizei statistice a rezultatelor, analiză grație căreia sunt selectați itemii reprezentativi pentru versiunea finală a instrumentului.

5) Validarea instrumentului

Itemii reprezentativi fiind selectați și versiunea finală a testului fiind constituită, urmează determinarea proprietăților metrice ale instrumentului. Proprietățile variază în funcție de natura instrumentului. În cazul instrumentelor normative se stabilesc normele care se reprezintă conform unei scale. În cazul instrumentelor criteriale se precizează scorul de referință [15]. Totodată, este important a sublinia că problema validității nu este doar de resortul autorului instrumentului respectiv. Validitatea nu este o calitate achiziționată odată pentru totdeauna. Orice inferență, pe care un utilizator vrea să o realizeze pornind de la rezultatele unui instrument, trebuie să fie obiectul unei validități specifice.

Referințe:

1. Laveault D., Gregoire J. Introduction aux theories testes en sciences humaines. - De Boeck universite, 1997.
2. Platon C. Evaluarea calității în învățământul universitar. - Chișinău, 2004.
3. Laveault D., Gregoire J. Introduction aux theories testes en sciences humaines. - De Boeck universite, 1997.
4. Molenaar N. Recent methodological studie son survey questioning quality and quantity, 1991, p.167-187.
5. Platon C. Evaluarea calității în învățământul universitar. - Chișinău, 2004.
6. Ibidem.
7. Dickes P. et al. La psychometrie. - Paris: PUF, 1994.
8. Stan A. Testul psihologic.- Iași: Polirom, 2002.
9. Cox E. The optimal number of responses alternatives for a scale // Jurnal of Marketing Research, 1980, no.17, p.407-422.
10. Schwajer K. The representational theory of measurement //Assessment Psychological Bulletin, 1991, no.110, p.618-626.
11. Cronbach L. Beyond the two disciplines of scientific psychology // American psychologist, 1975, vol.30, p.116-127.
12. Molenaar N. Recent methodological studie son survey questioning quality and quantity, 1991, p.167-187.
13. Haynes S. et al. Content validity in psychological assessment // Psychological Assessment, 1995, no.7, p.244.
14. Laveault D., Gregoire J. Introduction aux theories testes en sciences humaines. - De Boeck universite, 1997, p.83.
15. Ibidem.

Prezentat la 24.04.2012