

The typology of items used in computer assisted didactical evaluation

Prof.dr. Ion Gh. ROȘCA, Catedra de Informatică Economică, ASE București
Ec. Eugeniu CRISTESCU, Serviciul de Telecomunicații Speciale

The use of computers in educational evaluation systems requires the use of some specific item types. This paper presents evaluation techniques, used to elaborating knowledge tests and didactic ones, and suggests possible ways of categorizing them with direct applicability to the designing of didactic evaluation systems assisted by computer.

Keywords: summative evaluation, formative evaluation, criterion-referenced tests, norm-referenced tests, item, computer-based training system.

1 Introducere

Testarea academică este o parte fundamentală a procesului de predare-învățare. Aceasta este utilizată nu numai ca bază a ierarhizării studenților evaluați la încheierea procesului de predare-învățare, dar poate fi, și ar trebui să fie, folosită și la direcționarea predării, ca suport pentru dezvoltarea curriculum-ului, ca ajutor la evaluarea cerințelor și diferențelor, la evaluarea progresului și la motivarea învățării. Testele psihologice, testele de achiziție standardizate și testele realizate de profesor sunt utilizate frecvent pentru atingerea obiectivelor enunțate. Profesorului îi sunt necesare cunoștințe practice despre testare și despre construcția testelor pentru a putea selecționa și dezvolta teste care să măsoare corect rezultatele învățării. Aceste cerințe capătă un accent deosebit în contextul mai larg al evaluării realizate în cadrul sistemelor de instruire asistată de calculator (SIAC).

2. Evaluarea sumativă și/sau formativă ?

Măsurarea și evaluarea sunt folosite în educație din diferite motive. De exemplu, unii măsoară sau evaluează pentru: a ajuta la dirijarea studentului, pentru ierarhizarea rezultatelor sau acumulărilor studenților, pentru motivarea și realizarea instruirii acestora.

Nu toate instrumentele de măsură servesc la fel pentru scopuri diferite. Eficacitatea unei evaluări este deci dependentă de adecvarea instrumentelor de măsură folosite în procesul evaluării. De exemplu, evaluarea poate fi sumativă sau formativă.

Evaluarea sumativă cumulează rezultatele instruirii și progresele înregistrate de un grup.

Aceasta este tipul de evaluare utilizată la determinarea notelor, a aprecierii succesului instruirii și al eficacității curriculum-ului. Aceasta se efectuează la terminarea unui modul, curs sau curriculum.

Evaluarea formativă are loc pe parcursul modulului, cursului sau secvenței, în mod progresiv. Aceasta, din urmă, este folosită ca instrument de diagnostic, ca bază a corecțiilor și ca indicator al deficiențelor sau succeselor instruirii studenților. Acest indicator este folosit de formator, ca feedback imediat, pentru corectarea deficiențelor și multiplicarea rezultatelor pozitive, înainte de încheierea secvenței de instruire.

Evaluarea formativă este direcționată către zone reduse ale modulului, cursului sau curriculum-ului, în așa fel încât formatorul să poată realiza secvențe succesive de instruire individualizată, mult mai utile, astfel, fiecărui instruit în parte [Bloom et al., 1971]. Din aceste motive evaluarea sumativă este *normativă* (raportarea la normă), în timp ce evaluarea formativă este *criterială* (raportarea la criteriu). În primul caz, rezultatele obținute de instruiți sunt comparate între ele în vederea realizării unei ierarhii, iar în cel de-al doilea, rezultatele evaluării sunt comparate cu criterii sau standarde prestabilite.

3. Instrumente de evaluare criterială și/sau normativă ?

Testele sunt „probe obiective”, care se notează la fel de orice corector, întrucât răspunsurile date de cei evaluați se compară cu un „model de corectare” [Muster, 1970]. Testele au ca elemente constitutive item-urile.

Item-ul este cea mai mică entitate identifica-

bilă a unui test care este corelată cu unul sau mai multe obiective operaționale deduse, la rândul lor, dintr-un obiectiv didactic general [Roșca & Cristescu, 2004, p.33]. După Ebel testele criteriale diferă de cele normative [Ebel, 1979, p.11], astfel:

Testele criteriale:	Testele normative:
▪ sunt folosite pentru determinarea obiectivelor individuale realizate de instruiți;	▪ sunt folosite pentru determinarea nivelului general al cunoștințelor acumulate într-un anumit domeniu;
▪ indică realizarea de către instruit a unor obiective instrucționale specifice;	▪ indică poziționarea în cadrul unui grup;
▪ sunt focalizate pe un segment îngust de obiective;	▪ acoperă o arie largă de subiecte;
▪ sunt folosite pentru a determina dacă instruitul a realizat o sarcină specifică de învățare.	▪ sunt folosite pentru a indica gradul de succes în învățare al instruitului.

4. Tipuri de instrumente de evaluare

Instrumentele de evaluare (testele) diferă în funcție de scopul evaluării dar și de tipul acestora. Printre diferitele forme folosite, în mod uzual, de formatori sunt: testele standardizate, testele de tip obiectiv, testele de tip eseu, scalele de ierarhizare, listele de verificare, dispozitivele sociometrice, și testele de performanță.

5. Caracteristicile instrumentelor de evaluare

Indiferent de instrumentul de evaluare ales evaluatorul trebuie să fie preocupat de calitatea acestuia. În ciuda diversității instrumentelor de evaluare există câteva caracteristici de bază pe care acestea trebuie să le întrunească. Caracteristicile care asigură calitatea instrumentelor de evaluare sunt: *validitatea*, *fidelitatea* și *utilizabilitatea*.

5.1 Validitatea

Validitatea reprezintă gradul în care un instrument de evaluare măsoară ceea ce a fost destinat să măsoare.

Toate procedurile de determinare a validității unui test iau în considerare relația dintre performanța instrumentului de evaluare și faptele observabile relative la caracteristica evaluată. Procedurile care asigură validitatea pot fi

grupate în patru categorii:

- *validitatea de conținut*;
- *validitatea predictivă*;
- *validitatea concurentă*;
- *validitatea de construct*.

În literatura de specialitate este menționată și o a cincia categorie *validitatea de fațadă*. Aceasta nu se referă la validitate în sens tehnic. Pentru evaluat exprimă măsura în care testul este relevant și important [Stoica, 2003, p.45].

5.1.1 Validitatea de conținut

Validitatea de conținut se preocupă în primul rând cu fragmentarea în mod adecvat a părții de conținut sau comportament ce va fi supus evaluării. Aceasta este folosită în mod curent la testele de evaluare a achizițiilor și ar trebui utilizate de formatori pentru toate testele pe care le dezvoltă. Pentru asigurarea unui grad înalt al validității de conținut partea de conținuturi ce urmează a fi testată trebuie analizată cu atenție pentru a fi siguri că aspectele majore ale acestora sunt acoperite în mod adecvat. „Aprecierea validității de conținut se face de către experți, prin estimarea concordanței dintre itemii testului și obiectivele operaționale definite pe baza conținuturilor esențiale ale programului de instruire” (Ibidem, p.44).

5.1.2 Validitatea predictivă

Validitatea predictivă se referă la eficacitatea cu care un test prognozează nivelul succesorilor viitoare ale instruitului. Rezultatele testului sunt verificate comparativ cu performanțele ulterioare. Acest tip de validitate este folosit în mod frecvent la testele de la angajare, de diferențiere a candidaților la admitere etc.

5.1.3 Validitatea concurentă

Concordanța dintre rezultatul unui test cu cel obținut cu ajutorul altor criterii care măsoară succesul, în aproximativ același timp, se numește *validitate concurentă*. Acest tip de validitate este relevantă pentru testele folosite în scopul diagnozei stării existente.

5.1.4 Validitatea de construct

Validitatea de construct a unui test exprimă gradul de acuratețe până la care un test măsoară o caracteristică teoretică sau un anumit „construct”. Validarea construct-ului implică acumularea datelor din mai multe surse.

Printre factorii care influențează validitatea unui test (Ibidem, p.45) se pot menționa:

- indicații neclare;
- nivelul de dificultate necorespunzător al item-urilor;
- item-uri de calitate slabă;
- test prea scurt;
- administrarea și corectarea necorespunzătoare a testului;
- caracteristicile grupului cărui i se administrează testul.

5.2 Fidelitatea

Fidelitatea unui test se referă la constanța scorurilor obținute de aceleași persoane în situații diferite sau cu seturi diferite de item-uri echivalente. Determinarea fidelității unui test implică tehnica corelării a două seturi independente de rezultate. Corelația este exprimată sub forma *coeficientului de corelație* care exprimă gradul de similaritate dintre două seturi de rezultate (scoruri). În cazul unei relații perfect pozitive valoarea acestui coeficient este +1,00.

Factorii care afectează fidelitatea sunt sintetizați în patru grupe:

1. Factori care țin de construcția testelor:
 - lungimea testului;
 - claritatea instrucțiunilor;
 - caracteristicile fizice ale testului;
 - ambiguitatea item-urilor.
2. Factori care țin de punctarea testelor:
 - obiectivitatea testului;
 - claritatea cheilor.
3. Factori care țin de administrarea testelor:
 - condițiile de mediu (lumină, căldură, zgomot etc.);
 - timpul alocat testului;
 - indicațiile administratorului de test.
4. Factori care țin de studenți:
 - motivația;
 - atitudinile;
 - starea fizică;
 - emoțiile.

Cu cât un test include mai multe dintre aceste condiții cu atât mai mare va fi fiabilitatea acestuia. Încercând să răspundă la întrebarea: Ce este mai importantă, validitatea sau fide-

tatea testului ? profesorul Adrian Stoica observă că:

- „Un test poate fi fidel fără a fi valid, deoarece testul poate măsura altceva decât a fost desemnat să măsoare.
- Dacă un test nu este fidel, acesta nu este valid.
- Fidelitatea este o condiție necesară, dar nu și suficientă pentru validitate.
- Din punctual de vedere al evaluării în clasă, validității ar trebui să i se acorde o importanță mai mare.” [Stoica, 2003, p.46]

În anumite contexte apare un „conflict” între validitate și fidelitate: creșterea primei conduce la o scădere a celei de-a doua și invers. Această legătură este redată în Figura 1.

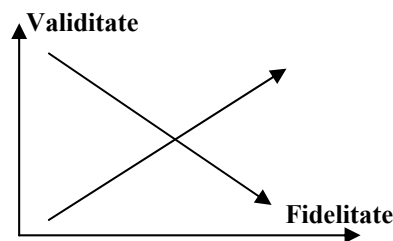


Fig. 1. Legătura validitate-fidelitate

5.3 Utilizabilitatea

Utilizabilitatea reprezintă gradul în care un test poate fi folosit cu minimum de timp, costuri și energie. Există multe teste standardizate care sunt neutilizabile în clasă de către formator deoarece sunt prea scumpe, prea greu de administrat sau prea dificil de interpretat.

Obiectivitatea este un ingredient al fidelității. La testare, obiectivitatea se referă la scăderea arbitrariului (de exemplu, „alegerile norocoase” sau „la întâmplare”) și a erorilor de notare.

6. Tipologia item-urilor

Luând în considerare drept criterii *obiectivitatea în notare* și *tipul de răspuns așteptat* item-urile se împart în categorii și subcategorii (Ibidem, p.54), astfel:

Criteriul 1: Obiectivitatea în notare

1. Item-uri obiective

- 1.1. Item-uri cu alegere duală
- 1.2. Item-uri cu alegere multiplă
- 1.3. Item-uri de tip pereche

<p>2. Item-uri semiobiective 2.1. Item-uri cu răspuns scurt 2.2. Întrebări structurate</p>
<p>3. Item-uri subiective 3.1. Item-uri cu alegere duală 3.2. Item-uri cu alegere multiplă 3.3. Item-uri de tip pereche</p>
<p> criteriul 2: Tipul de răspuns așteptat 1. Item-uri pentru care răspunsul este selectat sau marcat dintr-o listă de opțiuni 2. Item-uri pentru care răspunsul trebuie elaborat</p>

7. Tipologia item-urilor utilizate în cadrul Sistemelor de Instruire Asistată de Calculator (SIAC)

Elaborarea unui test de cunoștințe în variante asistate de calculator comportă o suită de item-uri de a căror folosire adecvată depinde calitatea testelor [Roșca et al., 2002, p.36]. Aceste tipuri sunt:

- Întrebări cu răspuns adevărat/fals (True/False Questions).
- Întrebări cu variante de răspuns multiple (Multiple Choice Questions).
- Întrebări cu spații libere de completat (Fill-in-the-Blank Questions).
- Întrebări de tipul uneori/întotdeauna/ niciodată (Sometime/Always/Never Questions).

Opțiunile de prezentare posibile sunt:

- o întrebare pe pagină;
- toate întrebările pe o pagină, cu întrebări grupate după tip (adevărat/fals, completare de spații libere, alegeri multiple etc.);
- toate întrebările pe o pagină, fără gruparea întrebărilor după tip;
- pagini separate pentru fiecare tip de întrebare.

Opțiunile de feedback pentru cursanți sunt:

- Corect sau incorect (*Correct or incorrect*);
- Oferirea unei recompense (*Provide reward*);
- Oferirea unei recompense cu o miză suplimentară (*Provide reward with additional challenge*);
- Furnizarea răspunsului corect (*Provide correct answer*);
- Sugerarea unui studiu pentru remediere

(*Suggested remedial study*);

- Furnizarea instruirii pentru remediere (*Provide remedial instruction*);

- Furnizarea scorului (*Provide score*);

- Combinații rezonabile ale acestor opțiuni.

Opțiuni de marcare a timpului consumat în timpul testării:

- După fiecare întrebare (*After each question*);

- După unele grupuri predefinite de întrebări (*After some predefined group of questions*);

- După încheierea testului (*After the test has been completed*).

O altă taxonomie [Jinga et al., 1999] împarte item-urile în două tipuri: de *tip închis* și de *tip deschis*, în diferite variante.

În cadrul SIAC este limitată utilizarea item-urilor de tip „eseu” sau item-urilor „de construit” deoarece răspunsurile subiecților supuși testării furnizează răspunsuri divergente (dar pertinente) care nu pot fi descrise algoritmic. Chiar și în practica pedagogică aprecierea acestora se face diferit, pe baza unor criterii și nu pe baza unor răspunsuri definite riguros.

Item-urile de tip „închis”

Acest tip de item are o aplicabilitate, cel mai adesea, sub forma *item-urilor cu alegere multiplă*. Structura acestora este formată din:

- **rădăcină** (întrebarea adresată subiectului testat);

- **răspunsurile preformulate**.

Cele din urmă cuprind la rândul lor *cheia* și *distractorul* (răspunsul/răspunsurile eronate). La acest tip de item se întâlnesc variantele:

a) *Item-ul cu răspuns binar*. Este necesară alegerea unuia dintre cele două răspunsuri preformulate.

b) *Item-ul cu cheia formată dintr-un singur răspuns preformat*.

c) *Item-ul cu răspuns multiplu*. Are cheia formată din mai multe răspunsuri preformulate.

d) *Item-ul cu răspuns grupat*. Are formulate atât cheia cât și distractorul din mai multe răspunsuri preformulate grupate.

e) *Item-ul de asociere tip cauză-efect*. Are rădăcina formată dintr-o frază cu două pro-

poziții. Subiectul testat trebuie să analizeze dacă propozițiile sunt adevărate și dacă există o relație cauzală între ele.

f) *Item-ul de tip pereche sau corespondență*. Necesită realizarea corespondenței dintre anumite elemente dispuse în două coloane.

g) *Item-ul de ordonare*. Necesită realizarea unei succesiuni fundamentată de o anumită logică a unor elemente, etape, obiecte etc. dispuse într-o ordine aleatoare.

Item-urile de tip „deschis”

Item-urile de tip deschis, utilizate în cadrul SIAC, au avantajul că anulează posibilitatea de ghicire a răspunsului corect, răspunsul solicitat trebuind să fie construit de către subiectul supus examinării. Proiectantul se confruntă, însă, cu dificultăți datorate caracterului plurisemantic al vocabularului și în același timp cu posibilitatea redusă de exprimare sub formă algoritmică a acestui tip de item. Aceasta conduce la restrângerea posibilității de cotare a răspunsurilor prin intermediul software-ului educațional și generează controverse privind aprecierea obiectivă a rezolvării acestora.

Cu toate aceste dificultăți, în practica SIAC s-au impus următoarele tipuri de item-uri de tip deschis:

a) *Item-ul de tip „propoziție lacunară”*. Necesită completarea unor spații goale în cadrul unei propoziții sau fraze;

b) *Item-ul de tip „desene lacunare”*;

c) *Item-ul de tip „desen de adnotat”*. Necesită precizarea anumitor părți lipsă ale desenului cuprins în item. Este o variantă a item-ului de tip „propoziție lacunară”.

Unele dintre tipurile menționate anterior se regăsesc în practica SIAC sub două forme (Op.cit., p.88-89):

1) cu editarea lacunei și construirea răspunsului de la tastatură;

2) cu editarea lacunei și alegerea răspunsului dintr-un meniu.

Utilizarea calculatorului în procesul de evaluare oferă posibilitatea evaluatorilor să realizeze item-uri care se prezintă în formă de tip „deschis” (prezintă una sau mai multe lacune) dar în conținut sunt de tip „închis” (cel evaluat primește informații din care poate să aleagă răspunsul corect). Informațiile oferite

depășesc numărul răspunsurilor corecte, aceasta solicitând, pe de o parte, un efort intelectual din partea celui examinat, iar pe de altă parte, posibilitatea de a ghici anumite răspunsuri prin alegerea lor la întâmplare.

8. Analiza de item-uri

Analizând răspunsurile date la test de către cei evaluați putem determina calitatea item-urilor, indiferent dacă acestea au fost elaborate de profesori sau de instituții specializate. Analiza se desfășoară atât după pretestare cât și după aplicarea testelor pe eșantioane mari de evaluați. Ea are ca scop găsirea răspunsurilor la următoarele întrebări (N. Gronlund, 1981, apud Stoica, 2003, p.92):

- „- Funcționează item-urile așa cum s-a intenționat?
- Au item-urile o dificultate corespunzătoare?
- discriminează item-urile testelor normative între elevii buni și cei slabi?
- măsoară eficient item-urile testelor criteriile rezultatelor învățării?
- Sunt distractorii eficienți (pentru item-urile cu alegere multiplă) ?”

8.1. Analiza de item-uri pentru testele normative

Dificultatea unui item este redată de numărul sau procentul elevilor care au răspuns corect la item-ul respectiv. *Indicele de dificultate* (p) [Stoica, 2003, p.92] se calculează după următoarea formulă:

$$p = \frac{\text{numarul elevilor care au raspuns corect la item}}{\text{numarul total de elevi}}$$

unde, $p \in [0; 1]$.

O valoare a indicelui apropiată de 1 indică un item ușor, în timp ce o valoare apropiată de 0 semnalează prezența unui item dificil. Valorile acceptabile pentru $p \in [0,20; 0,80]$, cu o valoare dezirabilă de 0,5–0,6 care asigură o dispersie maximă a rezultatelor.

La item-urile cu alegere multiplă este foarte importantă calcularea indicelui de dificultate pentru fiecare opțiune în parte. Un distractor cu un indice mai mic de 5% (0,05) este considerat nesatisfăcător și trebuie eliminat.

De regulă, pentru determinarea puterii de discriminare a unui item se pornește de la ipoteza că evaluații care au obținut un scor mare la întregul test ar trebui să obțină scoruri bune la fiecare item al testului. Pentru calcularea *indicelui de discriminare* (d) se parcurg următoarele etape (ib., pp.92-93):

- se aranjează în ordine crescătoare scorurile tuturor elevilor;
- se împart elevii în 5 grupe în raport cu scorurile obținute (primii 20%, următorii 20% ș.a.m.d.);
- se utilizează formula:

$$d = \frac{\text{primii 20\% care au răspuns corect} - \text{ultimii 20\% care au răspuns corect}}{1/5 \times \text{nr. total de elevi}}$$

Indicele de discriminare are valori cuprinse între +1 și -1. Valoarea $d = +1$ indică faptul că toți elevii buni au rezolvat item-ul corect și toți elevii slabi au dat răspunsuri greșite, iar o valoare $d = -1$ arată că toți elevii slabi au răspuns bine, în timp ce toți elevii buni au răspuns greșit la item-ul respectiv. În acest caz item-ul trebuie eliminat. Sintetic, principalele valori ale indicelui de discriminare sunt redate în tabelul următor:

Valoarea lui d	Consecințe
$d \geq 0,40$	Item-ul este bun
$0,30 \leq d \leq 0,39$	Item-ul necesită o mică revizuire
$0,20 \leq d \leq 0,29$	Item-ul necesită o foarte atentă revizuire
$d \leq 0,19$	Item-ul se elimină

Sursa: [ib., p.93]

8.2. Analiza de item-uri pentru testele criteriale

Obiectivul testelor criteriale este acela de a cuantifica la ce nivel se află performanțele celui evaluat în raport cu obiectivele educaționale. Astfel, formatorul constată *ce știe* instruitul și *ce știe să facă*. În acest caz, dificultatea unui item nu mai este prestabilită ea fiind specificată de obiectivul operațional pe care îl testează. Atâta timp cât rezultatele evaluatului nu sunt comparate cu cele ale grupului de referință indicele de discriminare nu are nici o semnificație.

În situația dată, este important să determinăm dacă testul evaluează programul de instruire și învățare. Pentru aceasta trebuie să comparăm rezultatele evaluărilor obținute la același test aplicat înainte și după desfășurarea programului de instruire.

Pentru testul aplicat *ante* sau *post* instruire se calculează *indicele* de „sensibilitate la efectele instruirii”, după formula:

$$s = \frac{R_A - R_B}{N}, \text{ unde :}$$

R_A = numărul evaluărilor care au răspuns corect la item după programul de instruire

R_B = numărul evaluărilor care au răspuns corect la item înaintea programului de instruire

N = numărul de elevi care au participat la ambele teste

Unde, $s \in [-1; +1]$ cu o valoare dezirabilă cât mai apropiată de 1.

9. Băncile de item-uri

Pentru a reduce efortul de realizare a testelor este necesar ca item-urile să fie organizate în bănci de item-uri. *Banca de item-uri* reprezintă o colecție organizată de item-uri clasificate și catalogate care pot fi accesate, respectând regulile de acces și procesare impuse de administratorul acesteia în vederea realizării de noi teste. Un avantaj important al băncilor de item-uri este acela că sunt utilizate în toate formele de evaluare: formativă, sumativă, internă, externă, inițială, finală ș.a. Importanța băncilor de item-uri rezultă din scopurile în care acestea sunt create (vezi și Wood, Sturnik, 1969, apud Stoica, 2001, p. 88):

- (i) „existența unor colecții de item-uri de bună calitate, pentru diferite discipline, care să fie utilizate atât în evaluările interne, cât și în cele externe;
- (ii) construirea de forme paralele de teste care pot fi aplicate pe grupuri de elevi la momente diferite, cu asigurarea compatibilității rezultatelor;
- (iii) elaborarea, cu mai mare ușurință, a testelor standardizate;
- (iv) familiarizarea cât mai multor profesori cu modul de proiectare a item-urilor, în scopul îmbunătățirii acțiunilor evaluative la clasă;
- (v) identificarea celor ce pot scrie item-uri de bună calitate.”

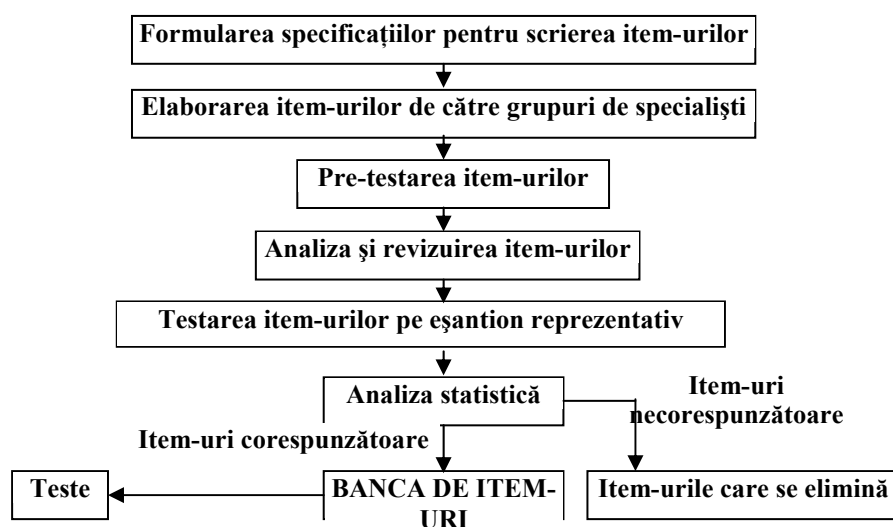
Clasificarea item-urilor în bănci se face în funcție de următoarele criterii:

- Clasă (nivel de vârstă)
- Disciplină
- Tip
- Obiectiv evaluat
- Date statistice (de exemplu, indicele de

dificultate și indicele de discriminare).

creare a unei bănci de item-uri (ibidem):

Schema următoare prezintă o modalitate de



În contextul existenței băncilor de item-uri, testele pot fi elaborate cu ușurință. Pe măsură ce aceste teste sunt administrate, banca este actualizată cu item-uri noi, urmând procedura descrisă anterior.

Concluzii

Considerăm că evaluarea asistată de calculator realizată în cadrul modulelor specializate din cadrul SIAC a devenit o realitate. În aceste condiții nu mai este posibil un management universitar eficient și performant care să facă abstracție de această modalitate digitală de evaluare. Articolul de față subliniază, în sinteză, câteva aspecte teoretice și practice privind taxonomia item-urilor, în general și a celor cu aplicabilitate în domeniul evaluării didactice asistate de calculator, în special. Aceste item-uri utilizate în practica sistemelor de evaluare didactică asistată de calculator fac obiectul implementării în cadrul modulelor specializate conținute în SIAC actuale.

Referințe bibliografice

📖 [Bloom et al., 1971] Bloom, B.S., Hastings, J.T., Madaus, G.F., *Handbook of Formative and Summative Evaluation of Student Learning*, New York: McGraw-Hill.

📖 [Cristescu, 2002] Cristescu Eugeniu: *Digitalizarea activităților de predare-învățare-evaluare trăsătură specifică pentru societatea informațională-societatea cunoașterii*, articol prezentat în cadrul Sesiunii de comunicări științifice cu tema: „Diferențierea și individualizarea instruirii – criterii, forme de organizare și

modalități de realizare“, Departamentul pentru pregătirea personalului didactic, ASE, 10 mai 2002, p.305-318, Editura ASE, București.

📖 [Cristescu, 2003] Cristescu Eugeniu: *Difficulties of didactic evaluation assisted by computer*, The proceedings of the 6th International Conference on Economic Informatics, May 8-11, Bucharest, Infocrec Printing House.

📖 [Ebel, 1979] Ebel, R. L.: *Essentials of Educational Measurement*, 3rd ed., Englewood Cliffs, New Jersey: Prentice-Hall.

📖 [Gronlund, 1981] Gronlund, N.: *Measurement and Evaluation in Teaching*, 4th edition, Macmillan.

📖 [Jinga et al., 1999] Jinga I., Petrescu A., Gavotă M., Ștefănescu V.: *Evaluarea performanțelor școlare*, Editura Aldin, București.

📖 [Muster, 1970] Muster Dumitru: *Verificarea progresului școlar prin teste docimologice*, E.D.P., București.

📖 [Roșca et al., 2000] Roșca Gh. Ion, Țăpuș Nicolae (coordonatori): *Internet și Intranet, Concepte și aplicații*, Editura Economică, București.

📖 [Roșca et al., 2002] Roșca Gh. I., Apostol C. G., Zamfir G., Bodea C. N.: *Informatica instruirii*, Editura Economică, București.

📖 [Roșca & Cristescu, 2004] Roșca, Gh. I., Cristescu, E.: *Îmbunătățiri metodologice privind realizarea item-urilor utilizate în evaluarea didactică asistată de calculator*, În: *Revista Informatica Economică*, nr. 1(29), București, Editura INFOREC.

📖 [Stoica, 2001] Stoica, Adrian: *Evaluarea curentă și examenul. Ghid pentru profesori*, Coordonator Adrian Stoica, Editura ProGnosis, București.

📖 [Stoica, 2003] Stoica, Adrian: *Evaluarea progresului școlar. De la teorie la practică*, Editura Humanitas Educațional, București.