

PUNCTAREA ITEMILOR ȘI CALCULUL SCORULUI TESTELOR DE AUTOR

Tudor BRAGARU

Universitatea de Stat din Moldova

Punctajul și gradul de complexitate al itemilor unui e-test de cunoștințe standardizat sunt ajustate empiric în urma unor experimentări pe eșantioane reprezentative. Punctarea și determinarea gradului de complexitate al itemilor unui e-test de autor sunt mai puțin elucidate în literatura pedagogică. În lucrare se prezintă o tentativă de punctare uniformă și determinare a gradului de complexitate pentru cele mai utilizate forme de itemi obiectivi și semiobiectivi prezente pe majoritatea platformelor moderne de e-learning/e-testing. Rezultatele unor așa teste se obțin prin simpla însumare a rezultatelor itemilor.

Cuvinte-cheie: item, punctare, grad de complexitate, e-test, scor e-test.

CALCULATING THE ITEMS' AND TEST'S SCORES OF AN AUTHOR'S E-TEST

Scoring and complexity of items in a standardized knowledge e-test are empirically adjusted based on the results of representative samples experiments. Approaches to scoring and determination of the degree of items complexity in an author's e-test are not sufficiently covered in pedagogical literature. This paper presents an attempt to uniformly score and determine the level of complexity of the most commonly used types of objective and semi-objective items present through the majority of the modern e-learning/e-testing platforms. The results of such tests are obtained by simple summing up of the items' results.

Keywords: item, scoring, level of complexity, e-test, e-test score.

Argument

În prezent, testarea pedagogică și-a modificat semnificativ abordarea într-o nouă direcție. Metodele moderne de e-testare tind să devină *deschise, dinamice și formative, cu feedback momentan și explicit, la nivel de variantă de răspuns, item, test*. Din metode exclusive de evaluare *e-testarea formativă, tematică și autoevaluarea* se transformă în *metode de învățare activă*. Cu ajutorul e-testelor subiecții nu doar asistă pasiv la performanțele lor, ci le pot autoevalua și dezvolta până la nivelul dorit. Utilizarea e-testelor formative și de autoevaluare permite identificarea imediată a unor lacune de studiu, *fără implicarea directă a profesorului*, dar care îi permite recuperarea, completarea și luarea unor decizii imediate de ameliorare a procesului didactic. Iar studentul obține posibilitatea de personificare a ritmului și traiectoriei sale de studiu, conform aptitudinilor și constrângerilor personale. e-Testele devin din ce în ce mai utilizate și în evaluarea/examinarea finală.

În condițiile *utilizării în masă a e-testelor*, cu elaborarea lor devin preocupăți din ce în ce mai mulți profesori din sistemul de învățământ de toate nivelurile, îndeosebi din învățământul universitar și continuu, unde se învață un număr mare de discipline dinamice. Însă, majoritatea profesorilor nu au o pregătire specială în domeniul testologiei și qualimetriei pedagogice.

O **problemă majoră** în elaborarea și desfășurarea e-testelor de autor este **lipsa recomandărilor privind punctarea uniformă** (termen echivalent cu **baremul de corectare**) a variantelor de răspuns ale itemilor de diferit tip și **atribuirea inițială a gradului de complexitate**. În general, nu este cunoscută o tehnologie de estimare a priori a caracteristicilor itemilor și a testelor de autor. Iar aplicarea directă a teoriilor clasice în acest sens este irațională, imposibilă (este complexă, costisitoare, de durată, necesită calificări speciale etc.). Majoritatea teoriilor clasice de punctare se referă, în temei, la teste omogene, care constau din itemi de aceeași formă și caracteristici, cu scoruri ajustate la nivel de test, cu determinarea empirică a gradelor de complexitate a itemilor, cu interpretarea normată a rezultatelor. Însă, în realitate, pentru testele de autor lucrurile stau exact invers: *itemii unui test bun sunt / ar trebui să fie de diferite forme, cu un număr diferit de elemente de răspuns, cu grade diferite de complexitate; cu feedback și notare momentane; cu interpretarea scorului final prin transformări liniare directe pe o scală numerică de notare*.

Astfel, devine incontestabilă **actualitatea identificării unor formule, tehnici simple de atribuire a gradului de complexitate, a scorurilor uniforme, comparabile la nivel de item/variantă de răspuns** pentru orice formă de item din care este alcătuit e-testul. Aceasta ar permite alcătuirea de teste de autor din itemi corelați între ei ca scor și complexitate cu posibilitatea interpretării directe a scorului final pe o scară de notare.

În lucrare este generalizată propria experiență de dezvoltare a itemilor și e-testelor de autor [Bragaru-2009, GIFT, 2010] prin prisma teoriilor clasice susținute de autori consacrați, ca [Albu-1993, Avanesov-IRT/RM/TM, Cucuș-2008, Radu-2008] și ținându-se cont de tendințele și posibilitățile oferite de platforme moderne de

e-learning, precum Moodle, Hot Potatoes, Claroline, AeL Siveco (AeL) etc. Utilizarea acestor recomandări ar permite creșterea calității și eficienței e-testelor de autor.

1. Tipuri și forme de itemi frecvent utilizate în e-teste

În toate lucrările de specialitate consultate [Avanesov-2005, Cucuș-2008, Radu-2008 etc.] itemii de evaluare sunt clasificați în paralel după două criterii: (1) *tipul de comportament solicitat pentru producerea răspunsului* – **itemi „închiși”**, care solicită selectarea răspunsului corect din mai multe variante propuse (precodificate), **itemi „deschiși”**, care solicită construirea unui răspuns și (2) *gradul de obiectivitate asigurat în notare* – **itemi obiectivi, semiobiectivi și subiectivi**.

Sumar, categoriile și formele de itemi sunt prezentate în *Tabelul 1*. Toate aceste forme de itemi sunt prezente pe majoritatea platformelor de e-learning/e-testing, dar cel mai frecvent utilizate sunt formele de itemi de tip **obiectiv** și **semiobiectiv**, dat fiind faptul că acestea pot fi corectate în mod automat, fără implicarea factorului uman.

Tabelul 1

Tipuri și forme de itemi prezente pe platforme de e-learning/e-testing

Categorie itemi	Obiectivi (închiși)	Semiobiectivi	Subiectivi (deschiși)
Forme de itemi	Itemi Binari (B) , Itemi Singulari (S)	Itemi de completare (<i>cuvinte lipsă</i> , L)	Rezolvare de probleme (P)
	Itemi cu alegere Multiplă (M)	Întrebări structurate/ Embedded (E)	Eseu structurat
	Itemi de (Co+) Ordonare (C/O)	Itemi cu răspuns liber scurt	Eseu liber (nestructurat)

Pe platformele moderne de e-learning/e-testing cel mai des sunt utilizați itemii obiectivi de tip **B, S, M, C** și semiobiectivi de tip **L, E** cu *notarea automată*. Itemii subiectivi sunt utilizați mai rar, preponderent pentru stocarea rezolvărilor și introducerea notelor atribuite de profesori în scopul evidenței și calculării automate a mediilor.

Itemii obiectivi, numiți și **închiși**, solicită identificarea răspunsului corect din mai multe variante propuse, respondentul nefiind pus în situația de a elabora răspunsul.

Itemii semiobiectivi solicită respondentului elaborarea unui răspuns scurt, completarea unui text lacunar sau răspunsul la întrebări/sarcini structurate. Răspunsul cerut este limitat ca formă, spațiu/loc prin natura sarcinii de lucru, care sugerează locul și numărul cuvintelor lipsă. Uneori itemii semiobiectivi sunt însoțiți de instrucții privind tipul răspunsului așteptat, modul de introducere a răspunsului, dacă se face diferențiere între litere mari/mici și alte elemente de sintaxă.

Principala calitate a itemilor închiși o constituie **obiectivitatea** în măsurarea/evaluarea rezultatelor învățării. *Aceste tipuri de itemi nu necesită implicarea directă a factorului uman în notare*, punctajul corespunzător acordându-se în funcție de marcarea elementelor răspunsului corect/incorect. Însă, pentru ca *sarcinile în formă de test* să se transforme în *itemi de evaluare* a rezultatelor învățării este necesară *stabilirea corectă a punctajului și scorului*, determinate de forma, conținutul, complexitatea itemului. Anume *itemii obiectivi constituie obiectul prezentei lucrări*. În e-teste itemii subiectivi sunt rar utilizați și în cadrul prezentei lucrări nu sunt examinați. Sumar, avantajele și limitările itemilor obiectivi și semiobiectivi sunt prezentate în *Tabelul 2*.

Tabelul 2

Avantaje și limitări caracteristice itemilor obiectivi și semiobiectivi

Avantaje	Limitări
<ol style="list-style-type: none"> 1. Sunt relativ ușor de construit și corectat. 2. Asigură o obiectivitate ridicată în evaluarea rezultatelor. 3. Pentru unele forme poate fi acordat punctaj parțial în funcție de corectitudinea răspunsului. 4. Permit evaluarea unui volum mare de rezultate ale învățării în timp scurt. 5. Corectarea și feedback-ul pot fi automate. 	<ol style="list-style-type: none"> 1. Măsoară rezultatele învățării situate la niveluri cognitive inferioare. 2. Încurajează o învățare bazată pe recunoaștere. 3. Nu pot fi folosiți pentru evaluarea unor rezultate de învățare complexă. 4. Pentru unele forme de itemi probabilitatea de „ghicire” este semnificativă. 5. Pentru itemi semiobiectivi este dificil a crea formulări absolut clare, este posibilă existența și altor răspunsuri corecte decât cele introduse de autor, sunt posibile ambiguități în interpretarea / sintaxa răspunsului.

Alcătuirea itemilor de o calitate superioară, corecți din punct de vedere tehnic și adecvați obiectivelor preconizate, constituie *cea mai laborioasă etapă a elaborării unui e-test* și nu e simplă deloc. După afirmația renumitului specialist V. Avanesov „...scrierea testelor de valoare nu este doar de gen științific, ci ține și de măiestrie... , de exemplu, poți să te înveți a așeza versuri, dar va rezulta oare din aceasta un poem?”. Unele recomandări generale privind *planificarea numărului de itemi, elaborarea diferitelor tipuri de itemi, includerea lor în teste* a se vedea și [Bragaru - 2009, 2010]. Ceea ce este mai puțin elucidat – modul de atribuire uniformă a punctajelor pentru diferite forme de itemi, examinat în continuare.

2. Strategii de punctare la nivel de item/variantă de răspuns

2.1. O analiză critică a punctării și calculului scorurilor conform teoriilor clasice

Căutările unor *formule și/sau recomandări de alocare uniformă a punctajelor la nivel de item/variantă de răspuns* pentru teste alcătuite din itemi de diferită formă și cu probabilitate diferită de ghicire a răspunsului nu s-au încununat cu succes.

Cea mai simplă tehnică de cotare a unui test de cunoștințe, propusă de teoriile clasice ca valabilă pentru toate tipurile de itemi, acordă *un punct fiecărui răspuns corect și zero puncte răspunsurilor greșite sau omise* [Albu - 1993]. În această abordare rezolvările itemilor sunt dihotomice: 1 – corect, 0 – incorect sau omis. Nu sunt aplicabili itemii de evaluare cu *răspunsuri parțial corecte*: cu alegere multiplă **M**, de (co-)ordonare **C** și semiobiectivi de tip **L** și **E**. Un alt neajuns este că punctajul maxim posibil este dat de numărul răspunsurilor corecte, fără a se ține cont de numărul distractorilor. Or, punctajul maxim nu este corelat nici cu probabilitatea de ghicire a răspunsului corect și nici cu efortul necesar pentru soluționare; același punctaj se atribuie unor itemi de dificultăți diferite. De exemplu, pentru itemi dihotomici tip **B** și **S** șansa de alegere aleatoare a unui răspuns corect dintre n variante $p = 1/n$. Or, pentru a obține un punct în itemi de tip **B** avem șansa de 1/2, sau 50%; dacă alegerea se face dintre 3 variante (tip **S**) – 30% etc., dintre 10 variante – 10%. Pentru comparație, probabilitatea obținerii unui punct în itemi de tip **C** din cinci perechi/elemente $p =$ numărul cazurilor favorabile/numărul cazurilor posibile $= 5/5! = 1/4! \approx 0,0416 \approx 4\%$. Adică, de aproximativ de la 2,5 (10%/4%) până la 12,5 (50%/4%) ori mai mică decât în cazul itemilor dihotomici (tip **B**, **S**). Pe de altă parte, probabilitatea obținerii unui punct într-un item de tip **M** cu patru elemente, dintre care două variante de răspuns corect și doi distractori, este de circa 83%. Agregarea directă a punctajelor în aceste exemple ar fi ca și cum am suma $1\text{ cm} + 1\text{ dm}$ și am obține 2. Deoarece măsurile sunt diferite (1 dm = 10 cm), agregarea corectă ar fi 11 cm sau 1,1 dm, în nici un caz însă 2. **Probabil, comentariile sunt de prisos.** Or, tehnica 0/1 este bună doar pentru *teste alcătuite din același tip de itemi, cu aceeași probabilitate de ghicire a răspunsului corect p*. În practică nu există așa cazuri, sau, dacă există, aceste cazuri sunt foarte rare. Deoarece teoria și practica pedagogică insistă asupra contrariului: **testele calitative trebuie să fie construite din itemi diferiți** ca formă, scor, complexitate, variante etc.

O formulă generală pentru un test compus din itemi cu alegere multiplă care ia în considerare probabilitatea de a ghici răspunsul corect a fost propusă de Traxler [Albu - 1993].

Conform acestei formule, se atribuie:

- 1 – un punct pentru fiecare răspuns corect;
- 0 – zero puncte dacă itemul nu a fost rezolvat/a fost omis;
- $k/(n-k)$ pentru fiecare răspuns greșit.

Cantitatea scăzută pentru un răspuns greșit $k/(n-k)$ este cu atât mai mare, cu cât raportul k/n , adică probabilitatea unui răspuns corect, este mai mare. Calculul scorului pentru așa test se efectuează după formula [adaptat după Albu – 1993]:

$$Scor = C - \frac{k}{n-k} G, \quad (1)$$

unde C = numărul de răspunsuri corecte date de cel examinat, iar G = numărul de răspunsuri greșite date de cel examinat.

Pentru cazul particular $k = 1$ testul constă doar din itemi cu răspuns singular și formula pentru calculul scorului se transformă în „*formula corectată pentru șansa succesului*” [Albu - 1993]. Evident, această formulă este aplicabilă doar pentru teste alcătuite din itemi cu aceeași probabilitate de ghicire a răspunsului corect, când toți itemii testului propun același număr de răspunsuri n și cer selectarea aceluiași număr de răspunsuri corecte k .

Încercările de a lua în considerare dificultatea itemilor atunci când se calculează scorul testului au dus la apariția unor noi teorii: (1) teoria răspunsului la item (Item Response Theory, IRT, [Deinego - 2009], [Avanesov-IRT]),

(2) Modelul Rush [Avanesov-RM] ș.a., care au la bază analiza post-test a răspunsurilor date de subiecți la itemi și sunt aplicabile preponderent în practica testelor standardizate și normalizate, elaborate de instituții specializate, fiind prea puțin utile pentru testele de autor (necesită cheltuieli mari, calificări speciale, durată mare de realizare etc.).

Concluziile de bază desprinse din succinta analiză a teoriilor clasice: (1) nici o metodă nu este totalmente potrivită pentru cazul testelor de autor, **nu oferă formule** de alocare uniformă **prin construcție** a punctajelor la nivel de item/variantă de răspuns pentru teste alcătuite din itemi de diferită formă și probabilitate de ghicire a răspunsului corect, (2) gradul de dificultate **se determină empiric**, necesită experimentări și procesarea statistică a datelor.

2.2. Abordări implementate în platformele moderne de e-learning

Majoritatea platformelor moderne de e-learning/e-testing (*Moodle, AeL, Hot Potatoes etc.*) utilizează *punctarea negativă* a răspunsurilor greșite și *punctarea diferențiată* a variantelor de răspuns. Strategia *punctării negative* a variantelor de răspuns este aplicabilă pentru itemi cu răspuns multiplu **M** și pentru itemi cu răspuns singular **S**. Utilizarea punctării negative la nivel de item *exclude necesitatea ajustării scorului final al testului, descurajează ghicirea răspunsurilor și simplifică calculul scorului prin simpla însumare a scorurilor itemilor*. Strategia *punctării diferențiate* a variantelor de răspuns poate fi aplicată pentru itemi cu **răspuns multiplu (M)**, itemi semiobiectivi de completare cu unul sau cu câteva cuvinte lipsă (**L**) și itemi subiectivi.

Problema acestor instrumente este că ele *nu propun și criterii/formule* concrete de atribuire uniformă a punctajelor la nivel de item în funcție de tip și număr de variante, lăsând autorii să decidă în mod voluntar. Însă, voluntarismul punctării itemilor este admisibil doar în cazuri excepționale a testelor construite din itemi de aceeași formă și caracteristici, care în practică lipsesc sau sunt foarte rar întâlnite. Or, ușurința de testare oferită de platformele de e-learning și e-testing, fără reglementările respective de construire a itemilor și testelor, duce la elaborarea pseudotestelor de autor cu efect contrar/opus scopului testării.

2.3. O schemă orientativă simplă de punctare și calcul al scorului

Având în vedere abordările clasice, pe de o parte, iar, pe de altă parte, posibilitățile platformelor moderne de e-learning, precum și necesitatea unor formule simple de punctare, precum și cerințele didacticii moderne de includere în examene/teste a itemilor cu, cel puțin, trei grade diferite de complexitate, în urma mai multor experimentări practice, se propune o *schemă simplă* de punctare uniformă a diferitelor tipuri de itemi în cadrul unui e-test de autor, cu *calcularea scorului testului prin simpla însumare a scorurilor itemilor*, așa cum sunt obișnuiți să facă profesorii în evaluările manuale:

$$Scor = \sum_{i=1}^m P(i) * V(i), \quad (2)$$

unde $V(i)$ este valoarea punctajului/scorului obținut la item, $P(i)$ – ponderea itemului în test, direct proporțională cu cantitatea de informații conținute și m – numărul de itemi în test. **În viziunea autorului, ponderea poate fi egalată cu 1 pentru toți itemii.** În primul rând, aceasta simplifică mult elaborarea testelor. În al doilea rând, conținutul testelor poate să difere, în funcție de tip/destinație. Or, ponderea conținuturilor verificate este mai comod de precizat în specificațiile testelor. A se vedea, de exemplu, [Bragaru-2009, 2010] sau matricea de specificare a testului, care rezolvă perfect această problemă. „*Matricea de specificații constă într-un tabel cu două intrări în care sunt precizate, pe de o parte, conținuturile care vor fi vizate și, pe de altă parte, nivelurile taxonomice la care se plasează obiectivele de evaluare*” [Matricea].

(1). Ținând cont de acestea, formula de calcul al scorului unui e-test de autor în schema propusă se reduce la simpla sumare:

$$Scor = \sum_{i=1}^m V(i) \quad (3)$$

Observație 1. Această formula este aplicabilă și la nivel de item, unde $V(i)$ este punctajul variantelor selectate i , iar m – numărul răspunsurilor date.

Dacă itemul este de tip **B, S, L**, scorul itemului este egal cu punctajul variantei selectate de răspuns. Dacă itemul este de tip **M, C(O), E**, fiecare răspuns va avea un punctaj pozitiv sau negativ, iar scorul total va fi egal cu suma punctajelor pentru toate răspunsurile selectate, ținându-se cont că la nivel de item **scorul se va egala cu zero, dacă suma va fi negativă** (vor fi selectate mai multe răspunsuri incorecte).

(2). În schema propusă punctarea se efectuează la nivel de item cu numere întregi, de exemplu, în diapazonul 1 – 10, astfel ca gradul de complexitate a itemului orientativ să fie egal cu scorul maxim, dar să fie precizat în funcție de forma itemului, probabilitatea ghicirii și a resurselor necesare pentru soluționare. Alternativele de răspuns se cotează ca parte a scorului atribuit la nivel de item. Astfel, pot fi definite 10 grade de complexitate, corespunzătoare punctajului itemului, cu excepția itemilor de tip *M*, *C* sau *O*, unde nivelul de complexitate poate fi precizat în funcție de probabilitatea ghicirii, timpul necesar soluționării și altele. Careva sugestii practice intuitive în acest sens conform experienței acumulate a se vedea în *Tabelele 3 și 4*. Dar, decizia rămâne la discreția autorului și potrivit regulamentelor interne. Itemii pot fi grupați în trei categorii, stabilite în majoritatea metodicilor didactice moderne: *itemi de complexitate mică – 1-3 puncte, medie – de la 4 la 7 și mare – de la 8 la 10 puncte*.

(3). La baza punctării uniforme stă **punctarea itemilor cu răspuns binar**, dat fiind că forma duală este cea mai elementară. Pentru **soluționarea corectă** a unui item binar vom acorda minimul posibil – **un punct**, iar **în caz contrar** (răspuns incorect sau omiterea itemului) – **zero puncte**. Respectiv, și scorul calculat la trecerea unui asemenea item va fi egal cu **unu** sau cu **zero**.

Observație 2. Exemplele ce urmează sunt itemi reali, pregătiți în șablonul GIFT (*a se vedea* [Bragaru-GIFT]). Cifra *x* între paranteze pătrate [*x*] semnifică punctajul itemului, iar între semnele :: se regăsește un cod flexibil și unicat al itemului în colecția și baza de itemi. În itemi ar trebui să se conțină obiectivele, dar ele nu sunt afișate, deoarece itemii sunt grupați în colecții nominative corespunzătoare obiectivelor. Cotarea în itemi cu răspuns singular și multiplu este indicată în procente față de scorul maximal. Dacă scorul maximal nu este indicat, atunci implicit el are valoarea unu [1].

Exemple de itemi binari:

::1B001:: Aria de acoperire este una dintre caracteristicile semnificative ale rețelelor informatice.

Feedback: DA, pentru că la scări diferite sunt folosite arhitecturi și tehnologii diferite.

::1B002:: Noua economie se reduce la economia digitală.

Feedback: NU, noua economie reprezintă o sinteză complexă între economia digitală, globalizare, inovare și dezvoltare durabilă.

Feedback-ul este posibil la nivel de item, varianta de răspuns și test integral și poate fi afișat sau nu, în funcție de necesitate, prin setările respective ale testului. În 1B001, 1B002 feedback-ul este la nivel de item. Din considerente de economie de spațiu în alți itemi feedback-ul nu va fi indicat, decât doar pentru ilustrarea diferitelor posibilități.

(4). **Punctarea itemilor cu cuvânt lipsă** se recomandă la nivelul itemilor binari (*un punct pentru răspunsul corect*) și o fracțiune de punct pentru răspunsuri parțial corecte. Uneori, acești itemi pot avea un punctaj și categoria de complexitate mai mare. *Exemple:*

[1]::1L001:: Rețeaua informatică globală a rețelelor lumii poartă numele de _____

- 100% Internet
- 90% internet

Feedback: Riguros exact (*internet = inter/între+network*) ≠ (*Internet = International+network*)

[1]::1L002:: Cel mai răspândit Internet serviciu este numit (abreviat) _____

- 100% Web
- 100% www

(5). **Punctarea și calculul scorului pentru itemi cu răspuns singular** poate fi:

$$Pct = Sc = \begin{cases} \pm [n/5] \\ 3 \leq n \leq 10, \end{cases} \quad (4)$$

unde *Pct* este punctajul atribuit variantei de răspuns corect (+) și incorect (-); *n* – numărul sumar de variante de răspuns în item, iar parantezele pătrate indică partea întreagă a câtului rotunjit. Condiția $3 \leq n$ semnifică apartenența la itemi de tip *S* (dacă $n = 2$, itemul se transformă în binar). Condiția $n \leq 10$ este dedusă din practică: în realitate itemi cu mai multe variante se întâlnesc destul de rar. Exemple de itemi cu răspuns singular:

[1]::1S001P1:: Cea mai ridicată fidelitate o au itemii:

- 100% Obiectivi
- -100% Semiobiectivi

- **-100%** Subiectivi

[2]::1S002P2:: Calitatea unui test de a măsura ceea ce își propune să măsoare se numește:

- **100%** Validitate de construct
- **-100%** Validitate de conținut
- **-100%** Fidelitate
- **-100%** Validitate de fațadă
- **-100%** Validitate criterială
- **-100%** Obiectivitate
- **-100%** Validitate predictivă
- **-100%** Validitate concurențială

Observație 3. Conform formulelor (4) și (2), itemii de tip *S* vor avea punctajul și scorul maximal **1 punct**, dacă $3 \leq n \leq 7$, sau **2 puncte** în caz contrar. Mai des acest tip de itemi este cotate cu un punct, este realizat pe platforme de e-learning ca radio-buton, cu posibilitatea de alegere doar a unui singur răspuns.

(6). Punctarea și calculul scorului itemilor cu răspuns multiplu. În virtutea formei sale, rezolvările itemilor cu răspuns multiplu pot fi parțiale, obținând un scor de la zero până la **Scor** conform condițiilor (1), observației (1) și formulei (3). Punctarea *Pct* a alternativelor de răspuns se va face conform formulei:

$$Pct = \begin{cases} +1, & \text{pentru fiecare raspuns corect} \\ -k/(n-k), & \text{pentru fiecare distractor} \\ 4 \leq n \leq 10, 2k \leq n \end{cases} \quad (5)$$

unde *k* este numărul răspunsurilor corecte; *n* – numărul sumar al elementelor de răspuns. Condiția $2k \leq n$ sugerează ca itemul să conțină, cel puțin, două variante corecte de răspuns și numărul lor să nu fie mai mare decât cel al distractorilor. Dacă $k = 1$ sau $n = 3$, itemul este/se poate ușor transforma în item de tip *S*. Dacă $n < 2k$, adică mai mult de jumătate din răspunsurile propuse sunt corecte, itemul ar trebui reformulat încât să ceară selectarea celorlalte $n-k$ răspunsuri din cele *n* propuse [Albu1 - 993].

Observație 4. Conform formulelor (5) și (2), scorul calculat la nivel de item de tip *M* va fi cuprins între zero (dacă se dau $n-k$ răspunsuri greșite) și maximal *k* puncte (dacă se dau *k* răspunsuri, toate corecte). Ori de câte ori se indică toate răspunsurile propuse, scorul itemului va fi nul sau mai mic ca zero, transformându-se în zero. Pentru $2k = n$ punctajul atribuit este cel mai mare $\lfloor n/2 \rfloor$. *Exemple:*

[2]::1M001P2:: Bifați numele protocoalelor de nivel IP în TCP/IP:

- **50%** IGMP

Feedback: Corect! Permite datagramelor IP să se răspândească în regim de grup

- **50%** ARP

Feedback: Corect! Mapează adrese MAC cu IP

- **-33.333%** HTTP

Feedback: Incorect! Este protocolul celei mai răspândite aplicații – Web

- **-33.333%** TFTP

Feedback: Incorect! TFTP, la fel ca și FTP, este un protocol de nivelul Aplicație

- **-33.333%** SMTP

Feedback. Incorect! Este protocolul aplicației de poșta electronică

[3]::1M002P3:: Pentru a putea fi aplicat, un test pedagogic trebuie să posede calitățile:

- **33.333%** Obiectivitate
- **33.333%** Fidelitate
- **33.333%** Validitate
- **-33.333%** Rigurozitate
- **-33.333%** Simplitate
- **-33.333%** Operaționalitate

Exemplul 1M001P2 conține feedback la nivelul fiecărei variante de răspuns, are punctajul egal cu doi, iar 1M002P3 cotate cu trei puncte nu are nici un feedback, decât scorul.

Pot exista excepții speciale, cotate special, în afara formulei (5), când un item cu răspuns singular S se poate masca sub un item de tip M , pentru a nu oferi implicit deducerea numărului de răspunsuri corecte pentru cei testați. Așa cazuri, foarte rare, nu au suport pe majoritatea platformelor de e-learning. Însă, dacă e necesar, ele pot fi modelate prin două elemente de răspuns corect sau doi distractori, de exemplu, cu valorile respective de $\pm 99\%$ și $\pm 1\%$. Evident, și gradul de dificultate în acești itemi poate fi special definit, în afara recomandărilor (2).

Șansa de a obține prin „ghicire” scorul maxim în itemi de tip M se calculează după formula:

$$Scor_max = 100\% \times C_n^k \times p^k \times q^{n-k}, \quad 10 \geq n \geq 4, \quad 2k \leq n \quad (6)$$

Pentru a ușura lucrul constructorului de itemi de tip M cu 4-10 elemente de răspuns se poate utiliza Tabelul 3. Toate variantele de răspuns corect și distractorii trebuie să fie plauzibile și să aibă punctaj egal între ele. Însă, în unele cazuri excepționale elementele de răspuns și distractorii pot avea punctaje diferite, dacă există necesități/indicații explicite în discriminarea lor. Cerința de suficiență pentru a preîntâmpina obținerea de puncte nemeritate prin bifarea tuturor elementelor de răspuns și a distractorilor este ca suma lor să aibă valoarea zero sau să fie negativă, ținându-se cont de faptul că suma negativă a punctajelor unui item se transformă în zero.

Tabelul 3

Recomandări de punctare uniformă a itemilor cu răspuns multiplu

Elemente de răspuns n	Inclusiv		Punctaj per răspuns corect	Punctaj per distractor	Scor maximal item	Șansa obținerii scorului maximal (%)	Grad de dificultate
	Corecte k	Distractori					
4	2	2	1	-1	2	37,5	1
5	2	3	1	-2/3	2	34,5	2
6	2	4	1	-2/4	2	32,9	3
	3	3	1	-1	3	31,3	3
7	2	5	1	-2/5	2	31,8	3
	3	4	1	-3/4	3	29,3	3
8	2	6	1	-2/6	2	31,1	3
	3	5	1	-3/5	3	28,2	3
	4	4	1	-1	4	27,3	4
9	2	7	1	-2/7	2	30,6	4
	3	6	1	-1/2	3	27,3	4
	4	5	1	-4/5	4	26,0	5
10	2	8	1	-1/4	2	30,2	4
	3	7	1	-3/7	3	26,7	5
	4	6	1	-2/3	4	25,1	5
	5	5	1	-1	5	31,6	4

Observație 5. Tabelul 3 pare dificil pentru utilizare. Însă, în realitate există multe instrumente de construire a itemilor (de ex., șablonul GIFT [Bragaru-GIFT]), care cer atribuirea punctajului la nivel de item (prin introducerea unui număr, orientativ între k și $[n/2]$), restul calculelor fiind executate automat, și oferind utilizatorului posibilitatea discriminării alternativelor prin corectări manuale.

(7). Punctarea și calculul scorului itemilor de asociere/ordonare a elementelor. Conform teoriilor clasice [Albu - 1993], punctajul pentru fiecare asociere corectă este același, mărimea lui fiind determinată ca cotă-parte din scorul/punctajul maxim al itemului, asocierile incorecte fiind punctate cu zero puncte. Pentru atribuirea uniformă a punctajului și calculul scorului pentru itemii de tip C pot fi considerați drept itemi cu alegere multiplă dintre m elemente cu k răspunsuri corecte, unde m este egal cu numărul perechilor care pot fi formate și k este numărul perechilor corect formate. Adică, punctajul maximal poate fi determinat ca $[n/2]$. Însă, comparativ, șansele de ghicire sunt mult mai mici pentru acești itemi, efortul și timpul necesar pentru soluționare fiind mai mari. Ca urmare, *potențialul de aplicare* a acestor itemi este mai înalt. Punctajul reco-

mandat pentru *itemii de asociere* cu diferite valori ale lui k în cazul $m = k$ este prezentat în Tabelul 4. Dacă $m > k$, probabilitatea ghicirii devine mai mică și punctajul ar putea fi mai mare.

Tabelul 4

Recomandări de punctare uniformă a itemilor-perechi ($m = k$) și de ordonare

Număr de perechi (k)	Total alternative de răspuns (k!)	Șansa de succes = $100\% * (k/k!) (%)$	Punctaj recomandat și grad de dificultate
3	6	20	1
4	24	16,7	2
5	120	4,2	3
6	720	0,83	4
7	5040	0,13	5
8	40320	0,02	6
9	362880	0,002	7
10	3628800	0,0003	8

Exemple:

[2]::C003P2:: Puneți în corespondență programele FP7 cu caracteristicile lor principale

Programul Cooperare

constituie elementul central al PC7

Programul Idei

reprezintă cercetarea pură, investigativă la frontierele științei și tehnologiei

Programul Oameni

se referă la îmbunătățirea carierelor europene în domeniul cercetării

Programul Capacități

semnifică construirea unei economii a cunoașterii

semnifică aplicarea biotehnologiei pentru sănătate

[2] [no_shuffle]::1O001P2:: Ordonați succesiunea etapelor/proceselor de bază ale unui proiect:

1

Inițierea

2

Planificarea

3

Execuția

4

Controlul

5

Punerea în funcțiune

(8). Punctarea itemilor structurați (*embedded*). Itemii structurați au un element comun, se referă la același fenomen, concept etc. Acești itemi evită influența altor tipuri de abilități, **concentrează evaluarea**, practic fiecare așa item poate fi utilizat ca test separat.

În Moodle itemii structurați pot fi construiți din combinarea a trei forme de itemi: cu alegere singulară **S**, multiplă **M** și cuvânt lipsă **L**. Ca urmare, itemii structurați întrunesc avantajele, dezavantajele și punctarea acestora. Scorul la item se calculează prin simpla însumare a scorurilor elementelor, ca și cum ar fi un test separat.

Exemplu de item integrat cu răspunsurile „deschise” introduse.

Pentru IP adresa 70.91.115.200/10 determinați:

1. Clasa IP adresei ✓

2. Masca implicită de rețea ✓

3. Masca IP adresei în format zecimal cu punct ✓

4. Scrieți formula și calculați: (în formule sunt utilizate semne de operație +/-/* de înmulțit/** exponenta/= egal și cifre/numere FĂRĂ SPAȚII între ele)

- 4.1. Numărul de biți rezervați pentru subrețea $10-8=2$ ✓
- 4.2. Numărul maximal de subrețele posibile $2^{**2}-2=2$ ✓ $10-8=2$
- 4.3. Numărul de biți rezervați pentru nod $32-10=22$ ✓
- 4.4. Numărul maximal de noduri posibile în fiecare subrețea $2^{**22}-2$ ✓
- 4.5. Pasul subrețelei $256-192=6$ ✓
- 4.6. Numărul subrețelei i , unde i este numărul de biți rezervați pentru subrețea $2^{*64}=124$ ✓
5. Identificatorul SUBREȚELEI i (în format zecimal cu punct) $70.128.0.0$ ✓
6. Valoarea binară și zecimală a măștii în octetul ce conține nr. de subrețea și o parte de nod (**despărțite prin punct, de ex: 1111100.252**) 11000000.192 ✓
7. Valoarea binară a octetului ce conține nr. de subrețea și o parte de nod 01011011 ✓
8. Identificatorul de REȚEA (în format zecimal cu punct) $70.0.0.0$ ✓
9. Identificatorul de SUBREȚEA (în format zecimal cu punct) $70.64.0.0$ ✓
10. Identificatorul de NOD a IP adresei inițiale (în format zecimal cu punct) $0.27.115.200$ ✓
11. Identificatorul primei subrețele atribuite cu primul nod $70.128.0.1$ ✓
12. Identificatorul primei subrețele atribuite cu ultimul nod $70.191.255.254$ ✓
13. Adresa de difuzare pentru prima subrețea atribuită $70.191.255.255$ ✓
14. Identificatorul ultimei subrețele atribuite cu primul nod $70.64.0.1$ ✓
15. Identificatorul ultimei subrețele atribuite cu ultimul nod $70.127.255.254$ ✓
16. Adresa de difuzare pentru ultima subrețea atribuită $70.127.255.255$ ✓

2.4. Transformările itemilor dintr-o formă în alta

Conform teoriilor clasice, pentru a îmbunătăți calitatea testării se utilizează așa-numita *analiză de item*, efectuată în urma unor experimentări ale testelor. De exemplu, se recomandă ca itemii la care au răspuns toți, la fel ca și cei la care nu a răspuns nimeni, să fie refăcuți sau excluși din colecții. La fel și distractorii trebuie revizuiți. Cu atât mai mult că platformele de e-learning/e-testing oferă instrumentele necesare de analiză. De asemenea, mulți autori constată că un item ușor (rezolvat de mai multe persoane) primește un punctaj mai mare decât un item dificil. Adică, se recomandă și modificările punctajului.

Toate acestea sunt incontestabile pentru testele sumative/finale. Însă, pentru testele de autor formative și de autoevaluare itemii și testele sunt valoroase nu atât pentru notare, cât pentru diagnoză, feedback, creșterea motivației și stimularea autoînvățării. Excluderea itemilor din colecții ar însemna și excluderea obiectivelor respective de învățare.

De aceea, în multe cazuri, în scopul îmbunătățirii lor și al păstrării raportului între obiective, conținuturi și evaluare, itemii pot fi transformați dintr-o formă în alta. De exemplu, itemii binari pot fi ușor transformați în itemi de completare sau în itemi cu răspuns singular, adăugând distractori. Fie că avem itemul:

::1B003:: Strategia Națională de edificare a societății informaționale „Moldova electronică” a fost adoptată în 2005 (true)

Acest item se poate transforma în item cu răspuns singular:

::1S003:: Strategia Națională de edificare a societății informaționale „Moldova electronică” a fost adoptată în:

- 100% 2005
- -100% 2004

- -100%2006
- -100%2007

Sau se poate transforma în item de completare:

::1L003:: Strategia Națională de edificare a societății informaționale „Moldova electronică” a fost adoptată în anul ____

- 2005

Simpla grupare a unor itemi binari cu răspuns fals și adevărat permite crearea de itemi cu răspuns singular sau multiplu. De exemplu, itemii binari:

::1B003:: e-Guvernarea înseamnă eliminarea în întregime a birourilor tradiționale (false)

::1B004:: e-Guvernarea trebuie să fie la dispoziția tuturor categoriilor de populație (true)

::1B005:: Siturile de e-Guvernare trebuie să fie publice (true)

::1B006:: e-Guvernarea se referă la folosirea TIC moderne pentru organizarea civică și politică a guvernului (true)

::1B007:: e-Guvernarea înseamnă participarea politica (e-Democrația) (false)

::1B008:: e-Guvernarea înseamnă incluziunea civică (e-incluziune) (false)

pot fi transformați într-un item cu răspuns multiplu:

1. [3]::1M003P3:: e-Guvernarea semnifică:

33.333% a fi la dispoziția tuturor categoriilor de populație

33.333% existența unor situri special destinate

33.333% folosirea TIC moderne pentru organizarea civică și politică a guvernului

-33.333% eliminarea în întregime a birourilor tradiționale

-33.333% participarea politică (e-democrația)

-33.333% incluziunea civică (e-incluziune)

Fără a intra în detalii, este evident că se pot efectua și transformări inverse.

Totodată, multe platforme (de exemplu, Moodle) permit generarea aleatoare a itemilor de asociere a k perechi, formate din itemii de tip L ai unei colecții. De exemplu, fie că avem o colecție de 30 de itemi de tip L :

::2L001:: Numărul maximal al rețelelor de clasa A

- 126

::2L006:: Numărul maximal de noduri într-o rețea de clasa C

- 254 etc.

Putem genera aleatoriu itemi de asociere ca combinații din n itemi luați câte k . Itemii rezultanți vor diferi prin elementele conținute. Numărul total de variante se calculează cu formula $C_n^k = n!/(k!(n-k)!)$. De exemplu, pentru $n = 30$ și $k = 5$ putem genera 142506 de itemi diferiți, 3003 – pentru $n = 15$ și $k = 5$. Astfel, potențialul acestui tip de itemi pentru testele formative și de autoevaluare este enorm. Deoarece la fiecare lansare a testului conținutul itemului va fi altul, se exclude copierea, nu e posibilă memorarea „mecanică” etc. Itemii generați aleatoriu arată în felul următor:

[2]::3CL01P2:: Pentru fiecare afirmație bifați finalul potrivit:

Numărul maximal al rețelelor de clasa A	126
Numărul maximal de noduri într-o rețea de clasa C	254
Numărul de biți pentru identificarea nodului într-o rețea de clasa A	24
Numărul de biți pentru identificarea rețelei de clasa A	8
Biți rezervați pentru identificarea rețelelor de clasa C	21

Un potențial enorm posedă și itemii parametrizați, în literatură fiind întâlniți sub numele de *itemi paraleli* sau *itemi-faset*. Dar, acest subiect merită a fi examinat separat.

Concluzii

Calitățile itemilor și ale testelor de autor trebuie asigurate preponderent prin construcție. Pentru a asigura uniformitatea atribuirii punctajului itemilor de diferită formă trebuie ținut cont de forma itemului, numărul de variante corecte și distractori, de „apropierea” lor, lungimea răspunsurilor, semnificația obiectivelor de referință etc. Pentru orientare în atribuirea inițială a punctajului și gradului de complexitate, în scopul evitării unor decizii total eronate cu efect anti-test, pot fi utilizate formulele și recomandările prezentate mai sus.

Însă aceasta în nici un caz nu exclude necesitatea analizei de item și desăvârșirea continuă a colecțiilor pentru administrarea testelor viitoarelor grupe de studenți, efectuată în baza statisticilor oferite de majoritatea platformelor moderne de e-learning/e-testing. De exemplu, platformele Moodle, AeL (ca și multe altele) au încorporate instrumente puternice de analiză statistică post-test a itemilor (*frecvența răspunsurilor, dispersia, coeficienții de discriminare etc.*), ceea ce permite desăvârșirea continuă a lor, inclusiv modificarea scorurilor inițial atribuite la nivel de element/variantă de răspuns sau item.

Este de menționat că punctarea rămâne, totuși, la discreția proiectanților de teste, care ar putea să se abată de la regulile recomandate, dacă se poate face o altă distincție clară cu privire la domeniu, destinație, scopul urmărit, performanța itemului etc. De exemplu, în medicină, strategia corectă de punctare a itemilor de tip **M**, **C** – componența unui medicament sau **O** – ordinea unor proceduri ș.a. este cea dihotomică: corect/incorect, adică nu există răspunsuri parțial corecte.

Concluzia de bază este că *nu pot exista rețete universale acceptabile de atribuire a scorului unui item*. Sau regulile sunt, în majoritate, confirmate prin existența unor excepții. Oricum, după cum afirmă ilustrul filosof Shopenhauer, „**Orice limitare ne face mai fericiți**”, ne ușurează alegerea, decizia, munca, ceea ce s-a și urmărit în această lucrare.

Bibliografie:

- [Albu-1993] Albu Monica, Pitariu Horia. Proiectarea testelor de cunoștințe și examenul asistat pe calculator. - Cluj-Napoca: Casa Cărții de Știință, 1993.
- [Avanesov-2005] Аванесов В.С. Форма тестовых заданий. - Москва: Центр тестирования, 2005.
- [Avanesov-IRT] Item Response Theory. Основные понятия и положения. <http://testolog.narod.ru/Theory59.html>, accesat la 14 mai 2012.
- [Avanesov-RM] Метрическая система Георга Раша (Rasch measurement, RM) <http://testolog.narod.ru/Theory68.html>, accesat la 14 mai 2012.
- [Avanesov-TM] Аванесов В.С. Теория и методика педагогических измерений. <http://testolog.narod.ru/Theory21.html>, accesat la 14 mai 2012.
- [Bragaru-2009] Bragaru T., Cîrhană V., Crăciun I. Testarea asistată de calculator. Metodologie. - Chișinău: CEP USM, 2009.
- [Bragaru-GIFT] Bragaru T. et al. Pregătirea itemilor de evaluare în MS Word și importarea în Moodle // Studia Universitatis (Chișinău, CEP USM), 2010, nr.2(32), p.87-92.
- [Bragaru-2010] Bragaru T. Aspecte metodico-didactice și bune practici de evaluare în Moodle. Modelare matematică, optimizare și tehnologii informaționale. Materialele Conferinței științifice internaționale. - Chișinău: ATIC, Evrica, 2010, p.328-346.
- [Chelyshkova-2002] Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. - Москва: Логос, 2002.
- [Cucos-2008] Cucos C. Teoria și metodologia evaluării. - Iași: Polirom, 2008.
- [Deinego-2009] Deinego N. Testarea adaptivă ca factor de optimizare a procesului de instruire în învățământul universitar. Autoreferatul tezei de dr., accesat la 19 mai 2012: <http://ebookbrowse.com/autoreferat-deinego-nona-pdf-d72889144>
- [Radu-2008] Radu I. Evaluarea în procesul didactic. - București: Editura Didactică și Pedagogică, 2008.
- [Matricea] Matricea de specificație și rolul său în proiectarea instrumentelor de evaluare. <http://www.scrigroup.com/didactica-pedagogie/Matricea-de-specificatie-si-ro25775.php>, accesat la 20 mai 2012.

Prezentat la 24.05.2012