

Strategii de evaluare

Lect. Cristian USCATU

Catedra de Informatica Economica, A.S.E. Bucuresti

Every "multiple-choice question" (MCQ) test bears the possibility that students will purely random guess some answers, thus obtaining an "undeserved score". The score will not reflect the student's real abilities and knowledge, instead being artificially raised. In some cases, totally unprepared students may get a score high enough to pass only through random guessing. There have been reports of students that used a dice to select the answer or simply selected the same answer (i.e. "B") for all questions and still got a pass score. In order to discourage/penalize guessing, different scoring strategies may be used. Their aim is to eliminate the undeserved score from the final score, more reliably reflect the student's knowledge and abilities, and produce a more accurate classification of students. The use of these strategies, as well as the strategies themselves, are a hot point of debate in the academic world. These strategies are known as "negative marking" or "formula scoring".

Key words: *negative marking, formula scoring, assessment, computer aided assessment, assessment-marking strategies, on-line testing, education, multiple-choice question.*

1 Introducere

Una din cele mai utilizate metode de evaluare în prezent este testul format din întrebări cu raspuns multiplu (MCQ), la care studentul trebuie sa aleaga raspunsul corect. Inerenta acestor teste este posibilitatea de a selecta raspunsurile la întâmplare. La un numar mare de întrebări, o parte dintre aceste raspunsuri vor fi corecte, aducând studentului un punctaj nemeritat, în cazul notarii clasice (un punct pentru raspuns corect, zero pentru raspuns gresit sau lipsa raspunsului). Acesta se adauga la punctajul meritat (corespunzator întrebărilor la care acesta a stiut raspunsul sau l-a dedus pe baza cunostintelor acumulate) oferind o imagine falsa asupra gradului sau de pregătire. Este posibil ca un student care nu s-a pregătit deloc sa aleaga la întâmplare toate raspunsurile si sa obtina totusi o nota de trecere. Sînt cunoscute cazuri de candidati care au folosit un zar pentru a selecta raspunsurile sau au ales aceeasi varianta pentru toate întrebările (de exemplu „B”) si au obtinut totusi nota de trecere.

În literatura de specialitate, punctajul obtinut astfel este numit „scorul maimutei” deoarece poate fi obtinut si de o maimuta care alege la întâmplare variantele de raspuns.

Pentru a descuraja ghicirea raspunsurilor si a penaliza astfel de metode, sînt folosite meto-

de de ajustare a notelor. Ele sînt cunoscute în general sub denumirea de „punctaje negative” sau „formule de ajustare a notei” si implica scaderea punctajului pentru fiecare raspuns gresit. Este normal ca în unele domenii raspunsurile gresite sa fie penalizate: de exemplu în medicina un raspuns gresit, datorat lipsei de pregătire sau unei pregătiri superficiale, poate însemna diferenta între viata si moarte pentru un pacient.

Aplicarea ajustării punctajelor depinde de tipul de test si de domeniul în care se aplica. În unele domenii este de dorit ca studentii sa încerce sa ghiceasca unele raspunsuri, folosindu-se de cunostintele do-bîndite (presupunînd ca nu sînt total straini de materie) în timp ce în altele (cazul medicinei de exemplu) se urmăreste eliminarea completa a acestei metode.

Strategiile de ajustare se pot aplica atît la nivelul fiecărei întrebări cît si la nivelul întregului test. Atunci cînd se aplica la nivel de întrebare, strategia presupune acordarea unui punctaj negativ pentru alegerea unui raspuns gresit. Cît de mare este acest punctaj (în valoare absoluta) depinde de tipul de strategie folosita. Atunci cînd se aplica o strategie la nivel de test, nu se acorda punctaje negative la fiecare întrebare, ci, din punctajul final se scade un numar de puncte determinat statis-

tic, ca o medie a punctelor obtinute prin ghicire. Modul de calculare a punctelor care se scad depinde de numarul de variante de raspuns si de punctajul acordat pentru un raspuns corect. În urma corectiei nu este afectata ierarhizarea studentilor ci se realizeaza o diferentiere mai clara a lor.

2. Strategii de punctare la nivel de întrebare

2.1. Punctaje negative. Pentru a descuraja „ghicirea” raspunsurilor, în locul metodei clasice, care acorda 1 punct pentru un raspuns corect si 0 puncte pentru un raspuns gresit, se poate aplica punctarea negativa a raspunsurilor gresite. Scopul este ca un student complet pregatit sa obtina un punctaj de 100%, în timp ce unul complet nepregatit sa obtina 0%. Ceilalti vor obtine un procent oarecare, reflexie a gradului lor de pregatire.

Se poate acorda un punct pentru un raspuns corect si -1 pentru un raspuns gresit. Pentru un test cu 2 variante de raspuns la fiecare întrebare, aceasta este o metoda convenabila de punctare: raspunzând numai la întâmplare, în medie 50% din raspunsuri sînt corecte iar 50% gresite. Punctajul final va fi 0. Pentru mai multe variante de raspuns însa, punctajul pentru raspunsurile gresite va fi mai mare în valoare absoluta decît cel pentru raspunsurile corecte. De aceea se pune problema balansarii punctelor pozitive si negative.

Daca s-ar acorda un punct pentru raspuns corect si -1 pentru raspuns gresit, la un test cu 5 variante de raspuns, numai prin ghicire se acumuleaza 20% din punctaj si se scade 80%, ajungînd la un total negativ (în loc de 0). Pentru a ajunge la un total nul, ar trebui sa se acorde 4 puncte pentru un raspuns corect si -1 pentru unul gresit. Acest mod de notare asigura respectarea ipotezei anterioare. Asemănator, pentru un raspuns corect se vor acorda 3 puncte la un test cu 4 variante, 2 puncte la un test cu 3 variante, 1 punct la un test cu 2 variante. În toate cazurile, pentru raspuns incorect se scade un punct.

Notînd cu C punctajul acordat pentru un raspuns corect, cu n numarul de variante de raspuns la fiecare întrebare (din care I este corecta si $n-I$ incorecte) si cu I punctajul nega-

tiv pentru fiecare raspuns incorect, avem $I = (\text{punctaj pentru un raspuns corect}) / (\text{numar de variante incorecte})$ adica $I = \frac{C}{n-1}$. I este

corectia aplicata la fiecare întrebare.

În toate variantele se acorda 0 (zero) puncte pentru lipsa raspunsului. Unele variante înclud explicit aceasta posibilitate, introducînd optiunea „Nu stiu”, a carei selectare nici nu mareste punctajul dar nici nu îl penalizeaza.

2.2. Permutari ale raspunsurilor. Se poate reduce foarte mult posibilitatea de ghicire a raspunsului daca întrebările au mai multe raspunsuri corecte, iar studentii trebuie sa aleaga permutarea corecta a acestora pentru a obtine punctajul aferent întrebării. Cu cît e mai mare numarul de elemente care trebuie permutate cu atît mai mica e posibilitatea de a ghici permutarea corecta (de exemplu, pe ntru 5 elemente numarul de combinatii posibile este $5! = 120$, deci probabilitatea de ghicire este de $1/120 = 0,84\%$).

În cazul acestei strategii de punctare, numai o pregatire temeinica asigura obtinerea punctelor. Ca varianta, se pot puncta si raspunsurile care sînt apropiate de cel corect (fara a fi exacte), cu o fractie din punctajul aferent întrebării. Dificultatea consta în aprecierea gradului de apropiere fata de raspunsul exact, stabilirea pragului peste care se acorda puncte si cuantificare a fractiei acordate.

2.3. Ordonarea preferintelor. Pentru acest tip de notare, fiecare întrebarea are un raspuns corect si mai multe gresite. Studentii trebuie sa ordoneze raspunsurile în ordine descrescatoare a preferintei. Punctele se acorda în functie de pozitie pe care a fost plasat raspunsul corect (de exemplu 4 pentru prima pozitie, 3 pentru a doua, 2 pentru a treia, 1 pentru a patra si 0 pentru a cincea pozitie). Punctajul final trebuie transformat în nota (de exemplu folosind una din metodele de mai jos).

O astfel de metoda de punctare este foarte utilizata la testele on-line: datorita posibilitatii de interactiune si raspunsului instant al calculatorului, se poate raspunde la fiecare întrebare pîna se da raspunsul corect. Puncta-

jul depinde de cîte încercari au fost necesare pentru gasirea raspunsului corect.

2.4. Siguranta raspunsului. Studentii trebuie sa asocieze raspunsului selectat un grad de încredere (de exemplu 1 pîna la 5). Punctajul aferent întrebării este gradul de încredere ales – daca raspunsul este corect sau negativul gradului de încredere asociat – daca raspunsul este incorect. Exista varianta ca punctarea negativa sa fie mult mai drastica (de exemplu dublul negativului gradului de încredere) în unele domenii. O astfel de strategie are sens în medicina, unde alegerea unui diagnostic gresit, cu un mare grad de încredere trebuie penalizata drastic.

2.5. Teste liberale. Metoda se aplica acolo unde se urmareste si punctare cunostintelor partiale. Studentul poate selecta mai multe raspunsuri la fiecare întrebare, punctajul dobîndit fiind o fractie din punctajul aferent întrebării, în functie de numarul de raspunsuri alese, daca între ele se afla si raspunsul corect. Altfel se acorda 0 puncte. Astfel, pentru o întrebare cu 5 variante de raspuns:

- Daca este selectat un singur raspuns, cel corect, se obtine întregul punctaj aferent întrebării ($4/4=100\%$, au fost eliminate 4 din 4 variante incorecte).
- Daca sînt alese 2 raspunsuri, se obtine $(4-1)/4=75\%$ din punctaj (în loc de 50% sanse de a obtine 0% sau 100% daca se alege un singur raspuns din cele 2).
- Daca sînt alese 3 raspunsuri, se obtine $(4-2)/4=50\%$ din punctaj (în loc de 33% sanse de a obtine punctajul întreg).
- Daca sînt alese 4 raspunsuri, se obtine $(4-1)/4=25\%$ din punctaj (în loc de 25% sanse de a obtine punctajul întreg).

Procentele de mai sus sînt valabile doar pentru cazul în care nu exista preferinte pentru raspunsurile alese (au preferinte egale), ceea ce în realitate se întîmpla destul de rar, dar se vede ca punctarea liberala este preferata celei clasice, acordînd credit si cunostintelor partiale.

3. Ajustare la nivel de test

3.1. Ajustarea punctajului. Pentru o mai mare usurinta în aplicare, corectarea punctajului se poate face la nivelul întregului test, în loc sa se faca la nivelul fiecărei întrebări, prin aplicarea punctajelor negative.

Folosind strategia de la punctul 2.1, se acorda un numar de puncte pentru fiecare raspuns corect, dar nu se scade nimic pentru raspunsurile gresite (se acorda 0 puncte). În acest fel, studentul obtine un punctaj (P) format din doua categorii de puncte: meritate (M – obtinute pe baza cunostintelor) si nemeritate (N – obtinute prin ghicire): $P=M+N$. Acest punctaj ar trebui corectat prin scaderea unui numar de puncte, aferente raspunsurilor incorecte (I), unde I este corectia pentru fiecare întrebare (calculata anterior) înmultita cu numarul de întrebări la care s-a raspuns gresit. Rezulta $P=M+N-I$.

Fie exemplul urmator: un test cu 100 de întrebări, 4 variante de raspuns la fiecare întrebare. Pentru raspunsurile corecte se acorda 3 puncte, iar pentru cele gresite se scade 1 punct. Sa presupunem ca studentul raspunde corect la 60 de întrebări (pe baza cunostintelor, obtinînd un punctaj meritat), alege un raspuns a întîmplare la 20 de întrebări (în medie 25% adica 5 vor fi corecte iar 15 incorecte – pentru care obtine un punctaj nemeritat, corectat prin aplicarea punctajelor negative), iar la alte 20 de întrebări elimina 2 variante (folosindu-se de cunostintele sale) si alege la întîmplare între celelalte 2 variante de raspuns; dintre acestea, în medie la 10 va raspunde corect iar la 10 incorect (punctaj partial – PP). Punctajul sau va fi format din:

$$M=60*3 = 180 \text{ puncte}$$

$$N=5*3-15*1 = 0 \text{ puncte}$$

$$PP=10*3-10*1 = 20 \text{ puncte}$$

rezulta $P = 200$ puncte.

Daca nu ar fi raspuns la cele 20 de întrebări la care a avut dubii între 2 variante, nu ar fi obtinut nici un punct pentru ele, punctajul final fiind 180, mai putin decît ar merita, tinînd cont de cunostintele partiale. Daca punctajele negative nu ar fi fost aplicate, studentul ar fi obtinut $180+15+30= 225$ puncte, mult mai mult decît cunostintele lui reale.

Se observa ca aceasta metoda de punctare încurajeaza aplicarea cunostintelor, chiar si partiale si descurajeaza „ghicirea” oarba a raspunsurilor. De asemenea este încurajata încercarea de a raspunde la toate întrebările, chiar si cele la care sînt dubii.

Punctajele obtinute în acest fel trebuie transformate în note. Transformare se poate face în mai multe feluri:

- Tinînd cont de domeniul posibil de punctare (-100 la 300 în exemplul anterior), 200 de puncte echivaleaza cu 75% din nota maxima deci nota 7,50. (7.00 daca nu ar fi raspuns la întrebările cu dubii, 8.13 daca nu se aplicau punctajele negative).
- Tinînd cont numai de punctajul maxim, 200 de puncte înseamna 66,7% din acesta adica nota 6,67. (6.00 daca nu ar fi raspuns la întrebările cu dubii, 7.50 daca nu se aplicau punctajele negative). Pentru studentii care au mai putin noroc decît media statistica, aceasta ar putea duce la punctaje negative, imposibil de transformat în note si cu efect psihologic negativ asupra lor.
- Tinînd cont numai de punctajul maxim, dar asimilînd toate punctajele negative cu 0. În acest fel studentul obtine tot 6.67, dar cei cu punctaje negative sînt asimilati notei 0.

3.2. Ajustarea (normalizarea) notei. În lipsa punctajelor negative, se poate aplica ajustarea direct la nota finala, pentru a elimina punctajul nemeritat. Daca N este nota neajustata, atunci nota ajustata A se calculeaza astfel:

$$A = \frac{n * N - 10}{n - 1}, \text{ unde } n \text{ este numarul de}$$

variante de raspuns al fiecare întrebare.

Aplicînd aceste corectii, nota de trecere se modifica. Daca nota de trecere ajustata este 5, la un test cu 5 variante de raspuns, studentul trebuie sa obtina nota 6 neajustata, la 4 variante trebuie sa obtina 6,25, la 3 variante 6,67, la 2 variante 7,5.

Se observa ca nota unui student perfect pregatit (10) nu este afectata de corectie. Un student perfect nepregatit, care obtine nota 0, prin ajustare ar obtine o nota negativa; calculele sînt însa facute pentru cazul mediu, cînd, numai prin ghicire, se obtin 20% din puncte, care în urma corectiei se reduc la 0%, o nota-

re consistenta cu situatia reala a studentului. Bineînțeles ca întotdeauna vor exista variatii de la medie, unii studenti vor ghici mai multe raspunsuri, altii mai putine. Se presupune însa ca studentii nu sînt total straini de materie si vor obtine un numar de puncte si pe baza cunostintelor acumulate. Notele negative pot fi asimilate cu 0, ca mai sus.

Exista si un alt efect interesant al aplicarii corectiei: cu cît un punctaj este mai mare, cu atît punctajul corectat va fi mai apropiat de el; cu cît este mai mic, cu atît va fi mai îndepartat punctajul corectat. În acest fel se face o distantare mai clara a studentilor si o diferentiere mai clara a gradului lor de pregatire.

3.3. Marirea notei de trecere. O alta strategie nu implica folosirea de punctaje negative ci reducerea importantei punctajului ghicit. Aceasta se poate realiza prin marirea punctajului de trecere si/sau marirea numarului de întrebări ale testului.

Pentru un test ale carui întrebări au 5 variante de raspuns, selectînd toate raspunsurile la întîmplare, în medie 20% vor fi corecte. În medie, pentru fiecare raspuns dat la întîmplare, studentul va obtine o cincime de punct. Daca punctajul de trecere „real” dorit este 50%, (la 50% din întrebări se raspunde corect pe baza pregatirii iar la celelalte la întîmplare) atunci punctajul „ghicit” este în medie $50\% * 20\% = 10\%$. Ca urmare se va stabili un punctaj de trecere de $50\% + 10\% = 60\%$ (nota 6, folosind punctarea conventionala). Daca întrebările testului ar avea doar 4 variante de raspuns, atunci punctajul „ghicit” ar fi de 25%, iar punctajul de trecere ar trebui ajustat cu $50\% * 25\% = 12,5\%$, devenind 62,5% (nota 6,25). Se observa ca marirea numarului de variante la fiecare întrebare duce la diminuarea posibilitatii de a acumula puncte prin ghicire, dar aici nu exista prea mult spatiu de manevra: 6 variante de raspuns sînt deja prea multe si duc la complicatii care fac mai mult rau decît bine iar cîstigul (din punct de vedere al acuratetei punctarii) obtinut este insignifiant. Se observa ca aceste calcule dau aceleasi rezultate ca în cazul aplicarii corectiei la nivel de test.

Marirea numarului de întrebări ale testului reduce probabilitatea de a obtine numai prin ghicire un punctaj de trecere. Conform [Mark Pettigrew, 2001], pentru un test cu 4 variante de raspuns la fiecare întrebare, numarul probabil de studenti care obtin punctajul de trecere (40%) numai prin ghicire variaza în functie de numarul de întrebări astfel:

Nr. întrebări	Procent de trecere
10	22
20	10
30	5
40	3
50	1.4
60	0.6
70	0

Se observa o reducere drastica atunci când se trece de la 10 la 40 de întrebări. Marirea numarului de întrebări are ca efect diminuarea împrastierii rezultatelor, acestea fiind mai apropiate de media teoretica (25% în acest exemplu). La un numar mic de întrebări, raspunsurile „ghicite” au o influenta mai mare asupra rezultatului final, în timp ce, la un numar mare de întrebări, aceasta influenta este aproape nula.

4. Posibilitati de folosire a calculatorului în testare

Natura testelor formate din întrebări cu raspuns multiplu le face usor utilizabile în sisteme automate de testare, unde punctajul se calculeaza dupa reguli clare si simple iar interventia evaluatorului uman nu este necesara. Utilizarea tehnicilor alternative de notare, care impun calcule (chiar daca sînt relativ simple) impune folosirea tehnicii de calcul, în special în cazul testarilor masive (sute sau mii de studenti).

Tehnologia informatiei, în special tehnologia web, se preteaza foarte bine la implementarea acestor teste. În prezent exista multe sisteme de testare care implementeaza un nivel oarecare de automatizare a evaluarii si chiar sisteme on-line, complet automatizate, majoritatea acestora implementînd sistemul traditional de punctare.

În cazul testelor on-line sînt posibile doua strategii: complet interactiva sau semi-

interactiva. În primul caz studentul primeste un punctajul pentru fiecare întrebare imediat ce a selectat raspunsul pe care îl considera corect. În al doilea caz punctajul va fi prezentat abia dupa ce studentul a raspuns la toate întrebările.

Pentru o strategie complet interactiva, în afara punctarii clasice, se poate aplica, de exemplu, tehnica „raspunde pîna cînd raspunsul e corect” (corespunzatoare tehnicii ordonarii preferintelor, descrisa mai sus). Toate celelalte tehnici sînt aplicabile, atît cele la nivel de întrebare cît si cele la nivel de test.

Automatizarea completa a testarii presupune, tipic, un server care furnizeaza testul si se ocupa de identificarea studentilor, colectarea raspunsurilor, evaluarea lor si calcularea punctajelor. Studentii se conecteaza prin terminale, folosind o interfata web (caz în care toate prelucrarile se fac la nivelul serverului) sau o aplicatie client special proiectata.

Deoarece testarea are loc, de obicei, simultan pentru toti studentii, se poate ridica problema numarului mare de terminale necesare simultan, situate în spatii adecvate testarii.. La un volum de sute sau mii de teste simultane (cazul examenelor de admitere), problema este de nerezolvat. Ca alternativa, testul poate fi sustinut pe serii, atunci cînd este disponibil un numar relativ mare de terminale iar numarul de serii este redus. Aceasta situatie este destul de rara în practica si ridica problema confidentialitatii testelor.

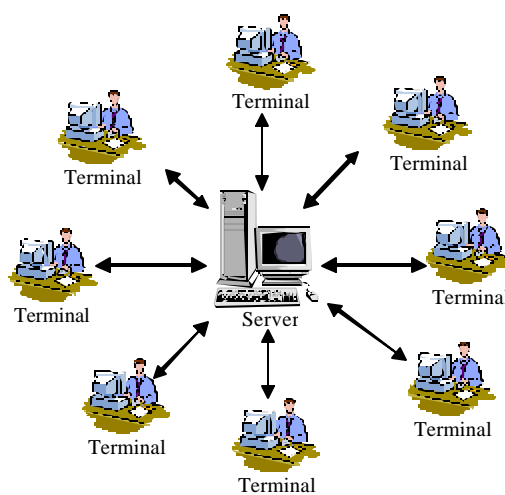


Fig.1. Structura unui sistem tipic de testare on-line

Folosirea de teste diferite pentru serii poate ridica o serie de obiectii din partea studentilor si nu numai. Alternativa cel mai des folosita este desfasurarea testului pe hîrtie, aceasta urmînd a fi „citita” si evaluata automat de un sistem format din scanner si calculator, care implementeaza o tehnica oarecare de punctare.

5. Concluzii

Testul format din întrebări cu răspuns multiplu este o metoda larg răspîndita de evaluare a cunostintelor. Principalele avantaje sînt usurinta de aplicare a lor, adaptabilitatea la testarea diferitelor aptitudini si niveluri de pregătire, acuratetea si obiectivitatea evaluării, posibilitatea de a verifica arii largi din materia studiată, posibilitatea de analiza complexă a răspunsurilor (în vederea îmbunătățirii sistemului educativ). Exista si multe argumente împotriva acestui tip de teste. Între acestea se afla dificultatea de a concepe întrebări „bune” (incluzînd consumul de timp), tendinta de a limita evaluarea la reamintirea unor date, posibilitatea ca, unii studenti, sa întrevada în textul întrebării mai mult decît a avut în minte autorul ei, numarul mare de întrebări necesare într-un test pentru a elimina efectul „ghicirii” răspunsurilor, dificultatea de a crea întrebări cu un grad adecvat de dificultate, nevoia de a crea mereu întrebări noi, pentru o materie care nu se schimba deloc sau prea puțin (pentru a contracara efectul „memorării” răspunsurilor) etc.

Tehnicile alternative de punctare ridica noi obiectii: aplicarea formulelor de calcul ridica probabilitatea de eroare la calcul (depasita însa prin folosirea tehnicii de calcul), distragerea atentiei candidatilor la întîlnirea unei noi tehnici de punctare (problema care se rezolva în timp), rezistenta psihologica a studentilor împotriva tehnicilor de punctare negativă.

Folosirea testelor cu răspuns multiplu si a schemelor alternative de punctare poate reflecta în mod adecvat (atît cît o poate face un test) nivelul de pregătire a studentilor si, mai mult, poate duce la ierarhizarea corectă a acestora.

Punctarea negativă este destul de puțin răspîndita în practica, aplicîndu-se în special în scolile medicale din Marea Britanie. Chiar si aici, utilizarea ei este controversată, în unele facultati fiind interzisă formal.

6. Bibliografie

- [Mark Pettigrew, 2001] Mark Pettigrew - *Random guessing on multiple choice questions*,
www.shu.ac.uk/services/lti/people/mp/mcq
- Martin Bush, Alternative Marking Schemes for On-Line Multiple-Choice Tests, www.caacentre.lboro.ac.uk/dldocs/bushmark.pdf
- Sheffield Hallam University, Learning and Teaching Institute, *Computer Assisted Assessment – practical guide*, 2003
www.shu.ac.uk/services/lc/cmeweb/grant/ltiweb/08/caapracguide.html
- Sheffield Hallam University, School of Education, *A Handbook to Support Teaching and Activity 2001/2002, MALT (MA in Learning and Teaching)*,
<http://www.shu.ac.uk/services/lc/cmeweb/grant/ltiweb/08/planbk20012002.doc>
- University of Leicester, *The Castle Toolkit – Designing and managing MCQ's, chapter 4 Score and Statistics*, 2002,
<http://www.le.ac.uk/castle/resources/mcqman/mcqcont.html>,
www.uct.ac.za/projects/cbe/mcqman/mcqchp4.html
- Learning and teaching support, LTSN-01 frequently asked questions, 2003, www.ltsn-01.ac.uk/resources/faq/index.html?my_num=24
- Madeleine Freewood, Negative Marking, 2001,
<http://www.shu.ac.uk/services/lti/resources/caapacticalguidance/negativemarking.html>
- Centre for Interactive Assessment Development, Don Mackenzie, Dave O'Hare, Chris O'Reilly, Helen Wilkins, *Assessment marking strategies*,
www.derby.ac.uk/ciad/dev/logical.htm
- A.R. Gardner-Medwin, Rational and Irrational Marking Schemes, presentation to Physiological Society, Cardiff, 1998,
www.ucl.ac.uk/~ucgbarg/tea/negmark2.htm

