Applied Medical Informatics
Original Research Vol. 31, No. 3/2012, pp: 69-76

Computer Aided Concept for Enhancing Motivation and Thinking while Teaching and Learning Medical Physiology

Mihail Gh. GLIGA [1,*], and Marius Şt. MĂRUŞTERI [2]

[1] University of Medicine and Pharmacy Tîrgu Mureş, Physiology Department, 38 Gh. Marinescu, 540000 Tîrgu Mureş, Romania.
[2] University of Medicine and Pharmacy Tîrgu Mureş, Medical Informatics Department, 38 Gh. Marinescu, 540000 Tîrgu Mureş, Romania.
E-mails: mihaigliga@yahoo.com; msmarusteri@yahoo.com

* Author to whom correspondence should be addressed; Tel.: +40745520208

Received: 1 September 2012 /Accepted: 29 September 2012 / Published online: 30 September 2012

Abstract
We wanted to refine the old computer aided concept we developed several years ago for teaching and learning medical physiology, by adding new tools to enhance motivation and thinking of both medical students and teachers. Since all the future exams of our University will rely mainly on a written, or even better, computerized multiple choice format, we focused on tools that motivate teachers in writing good multiple-choice questions (MCQ), not only for final exam but for teaching students as well, and tools to motivate students to perform MCQ tests during semester. We developed a concept of a Multifunctional or Enhanced MCQ which offer teachers, when using our software, the possibility to write questions with one statement and up to 10 answers which in their turn can be individually enabled or disabled by them. We conceived a precise protocol for calculation of Current P index and Current D index of each MCQ which is useful for a permanent, during semester, refining of poor quality questions or to adjust current overall difficulty of the MCQ set according to student performances and/or teacher expectancies.

Keywords: Computer aided concept; Motivation and thinking; Teaching and learning; Medical physiology; Cognitive domain.

Introduction

Medical physiology education on cognitive domain is performed in our University mainly by lectured courses in a very limited interactive way. We haven't implemented yet concepts like using computers to establish interaction between students and teachers during lectured courses [1] and we are well behind what was prefigured twelve years ago about implication of computers in teaching [2]. However, we had preoccupations in developing software tools and computerized systems to help the educational process [3,4]. These preoccupations proved to be useful since recently the leadership of our University decided to replace the end of semester oral exams with written exams and most of those written exams will consist probably on the evaluation of students using a MCQ set. We can develop our own computer aided concept for teaching, learning and assessing medical physiology and/or we can join to a worldwide knowledge database [5] and/or we can use readily concepts for improving quality of written assessment [6]. Analysis of quality of a MCQ set is linked to a given group of students who performed a test on that MCQ set and is done first by calculating for each MCQ the item difficulty P index and item discrimination D index. Item

P index is the ratio between the numbers of correct responses to total number of responses for that test item. For calculation of item D index one has first to order the given group of students by their overall scores on that MCQ set test. Then, for each item, D index is equal to item P index accordingly to the best 27% scorers minus item P index accordingly to the poorest 27% scorers [7]. This analysis is usually done only after a set of MCQs were used in testing on a group of students, so improvement of the quality of that MCQ set will be of benefit only for the next group/generation of students. More, the MCQ set used for students' final evaluation cannot be used for teaching during semester because students will tend to memorize the questions, no meter how many MCQs are in that set. On another side, students would be motivated more to test themselves on a MCQ set if they would know that same issues, from a given curricula, and same levels of knowledge, according to a taxonomy [8], will be expected from them at the final exam as those that MCQ set deals with.

We intended two things: first to develop a new concept of Current analysis of quality of a MCQ immediately after the Teacher wrote that MCQ and the students are able to perform tests on it and second to develop new functional MCQ types that will link strongly the teaching and learning process with the session official exam. Our aim was to develop tools that motivate teachers in writing good MCQ, not only for final exam but for teaching students as well, and tools to motivate students to perform MCQ tests during semester.

Material and Method

We used a set of MCQ which has been developed and was used in our University 15 years ago in order to evaluate candidates in Biology exam. This set consisted in 1962 questions/items classically structured in a statement and 5 answers. Items were grouped into 11 chapters accordingly to given Biology curricula. Medical physiology curricula on which we are teaching students is in fact an enhancement of Biology curricula used in previously mentioned exam. Each chapter-group of MCQ was further divided into 2 smaller chapters according to MCQ with 1 correct answer and MCQ with 2 correct answers. In the attempt for improvement of our database management system and our testing protocols we evaluated, using our old software, a group of 43 volunteered students over a period of one year. Every student could choose the moment when to join or leave the group, the frequency of testing and the moment when to perform the test. Every student could chose the curricular area from which he wanted to be tested and the number of questions to include in the current test and finally any student could chose the time allocated to finish the test. Our old version of software allowed the teacher to perform surveillance on students choices and consequently to guide work.

We analyzed first with our old protocol the quality of the mentioned MCQ set. In the old version of our software we focused on the student side mainly, that is we wrote computerized protocols that memorize for each student which question, out of those 1962 from the MCQ set, the student answered wrong and how many times. So the more times the student is performing tests, the more accurately our software can build a top of the most difficult MCQ for him and for every student apart. Our old version of software can perform surveillance of the current coverage of curricular area by each student by counting the number of times each student's test contained one or more MCQ from an identified curricular area ie increasing the counter for that area by 1/current test, without regarding how many MCQ were selected from that area for the current test. The old software doesn't provide the exact total number of responses for each MCQ and this is why we roughly calculated item P index as 1 minus the exact number of times the students answered wrong a question divided to the approximate total number of responses for that test item. We applied in the end same correction coefficient to all MCQ in order to fit P indexes between 0 and 1. Then we made an approximate calculation of D index too as, according to its definition, D index is the difference between two P indexes. We searched how strong is, in these conditions, the correlation between item P index and item D index by calculating the correlation coefficient [9]. We wrote a new protocol for a more precise calculation of an item P and D index, and the calculation

can be done immediately after the Teacher wrote that item and the students are able to perform tests on it.

We developed a concept of a Multifunctional or Enhanced MCQ based on the fact that motivation is in general defined as valence multiply by expectancy [10]. According to this definition, the student's motivation to deal during the semester with a question is the product of how important he feels that question is at that moment for him and the trust that he can answer easily to it. We offer to the teacher, by using our software, the possibility to write questions with one statement and up to 10 answers. The teacher can mark anytime from 5 to 10 answers of an Enhanced MCQ as active. When a student will perform a test on that EMCQ, the computer will extract randomly only 5 of all active answers but the statement will remain all the time the same. The teacher can split the answers into two parts as follows: the first 5 answers to test the lower 'recall' cognitive level on that knowledge the statement refers to, and the next 5 answers to test the higher 'understanding' cognitive level. The teacher may enable during the semester only the 5 'recall' answers of a statement and for the final exam the teacher may enable only the 5 'understanding' answers or all 10 answers or any combination of 5 to 10 active answers.

Writing text in a EMCQ, as simply as it looks, requires besides preoccupation, a lot of work. We used the statement in Figure 3 as a simple question we put in 2009 to 321 medical students at the beginning of year 2 and we expected for short written answers. We did it again to a similar number of students next year. Then we gathered the first 5 most frequent wrong answers and they became the first 5 answers in the MCQ.

Results

Data collected by our database management system while testing 43 students over a set of 1962 MCQ during a one year period and using protocols presented in material and methods paragraph, are like those in Table 1. One can note that student 30 performed 33 tests mainly from the entire curricular area (total curricular area coverage = 426) and have good results (total wrong answers= 192). Student 30 has overall much better results than student 16 who performed more tests (37) but with less curricular area coverage (252) and worse results (total wrong answers=1720).

Table 1. Examples of data acquisition while students are performing MCQ tests. Where 'not sel.' means not selected.

MCQ ↓	Current number of times a student answered wrongly the MCQ he tried to answer					Current number of times student selected for test the small chapter to which MCQ belongs					Small Chapter Number ↓
	Stud. 4	Stud. 10	Stud. 16	Stud. 30	Stud. 36	Stud. 4	Stud. 10	Stud. 16	Stud. 30	Stud. 36	
...	...	...	...	...	...	...	...	...	...	...	1
14	1	0	1	1	1	13	7	12	17	3	1
363	3	0	1	not sel.	1	12	10	12	20	4	4
845	3	1	2	1	not sel.	7	10	12	20	not sel.	10
1276	1	3	2	1	1	5	7	11	20	2	12
1951	3	0	1	not sel.	1	7	8	12	17	3	22
...	...	...	...	...	...	...	...	...	...	...	22
Total	1529	234	1720	192	206	152	158	252	426	44	Total
Total number of tests performed by 43 students/1year						36	19	37	33	7	939

A MCQ difficulty chart for the set of 1962 questions obtained after a total of 939 random tests performed by the 43 students in the group evaluated over a period of one year is shown in Figure 1. The teacher can take a close look at the most difficult items and refine the text in them if necessary.

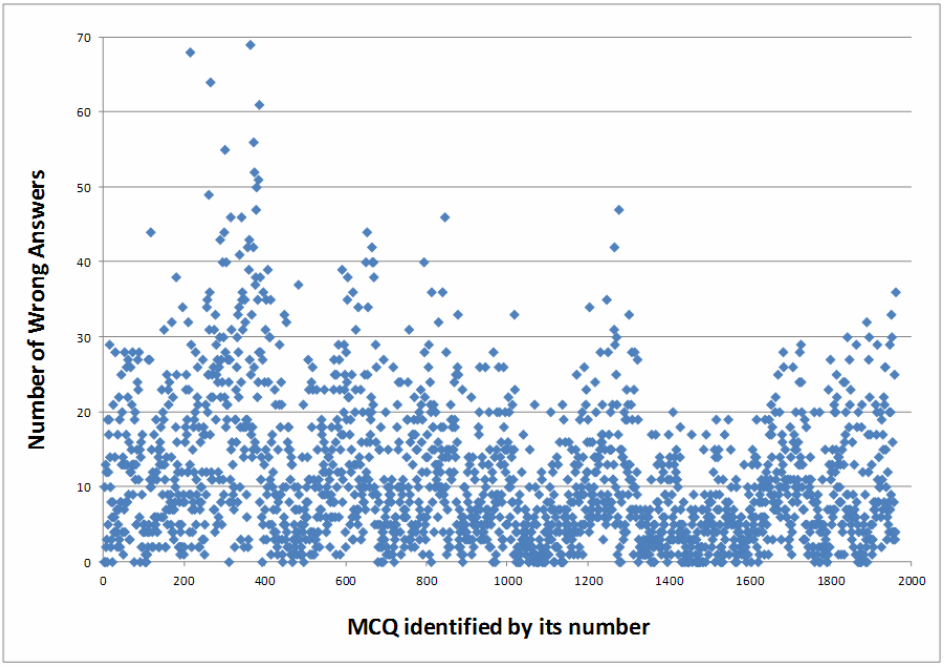

Figure 1. Difficulty of 1962 MCQs given by number of wrong answers when performed by 43 students over 1 year interval of time: P_Absolute

The teacher has information of the whole group so the teacher may adjust expectations accordingly to group evolution or, even better, may adjust education in order to achieve some previously established expectations. Note also that every student possesses and has access to a similar 1962 MCQ ranked, but individually shaped map. Every student is able to generate self tests comprising the most difficult personal questions ie those he answered wrongly the most number of times.

A rough calculation of P index and D index for each MCQ using data in Table I followed by a scatter plot of each item D index and P index leads to Figure 2.

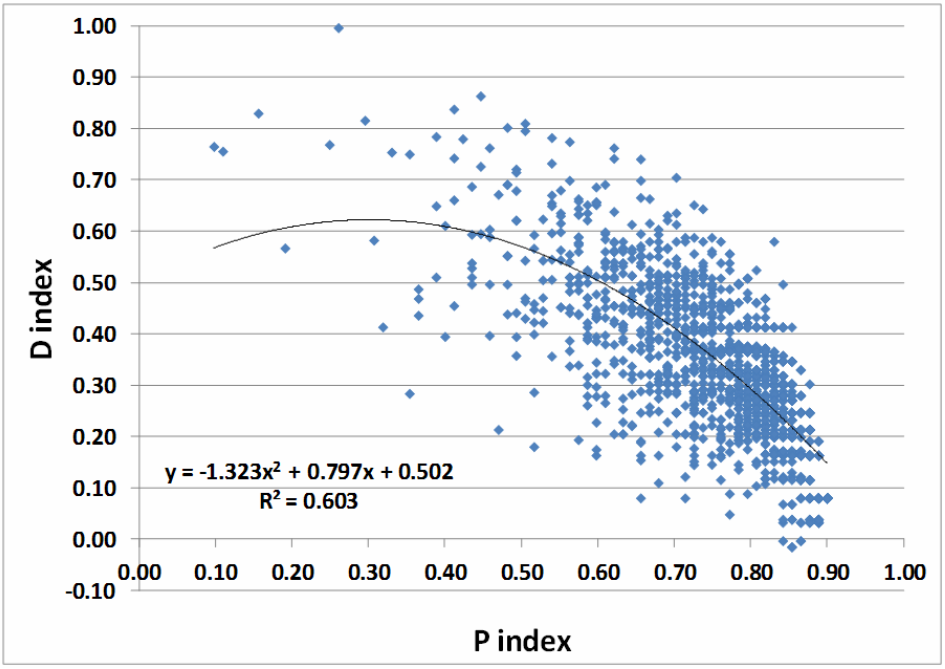

Figure 2. Correlation of rough D index as function of rough P index

Using a linear regression equation of rough D index as a function of P index we obtained a correlation coefficient r=0.752. When we used a second order model to describe the link between rough D and P indexes we obtained a correlation coefficient r=0.776. The correspondent coefficients of determination r^2 or R^2 are 0.567, respectively 0,603. The regression equation of rough D index as a function of P index, together with its graphical equivalent, is shown in Figure 2.

Nevertheless we conceived a more precise/refined protocol for calculation of Current P index and Current D index of each MCQ and here we present this refined protocol in Table 2 and Table 3.

Table 2. Data needed for a precise calculation of an item's current P index and D index

Data acquisition	Current number of times student answered wrong the MCQ he tried to answer					Current number of times student answered at all the MCQ as part of a test				
	Student 1	...	Student j	...	Student m	Student 1	...	Student j	...	Student m
MCQ 1	a	...	d	...	t	A	...	D	...	T
...	...	...	...	...	...	...	...	...	...	...
MCQ i	f	...	h	...	p	F	...	H	...	P
...	...	...	...	...	...	...	...	...	...	...
MCQ n	s	...	u	...	y	S	...	U	...	Y

Legend: In our work m=43 and n=1962. Otherwise any letter in this table can be any number, just that a, d, t, f, h, p, s, u, y are each less or equal then their corresponding capital letters respectively.

Current D indexes are useful for a permanent, during semester, refining of poor quality questions, ie those with a low D index, according to its definition. Current P indexes are useful for refining the current overall difficulty of the MCQ set to fit PSET index in order to be neither very easy, nor very difficult. Current overall difficulty of the MCQ set can be refined also according to student performances and/or teacher expectancies. At the end of a semester these P and D indexes will become fixed and a thoroughly final pre exam MCQ analysis of the set can be performed.

Table 3. Protocol for calculation of item's current P index

Calculation of ratio of each MCQ and for each Student						Calculation of Current Difficulty P index for MCQ
	Stud. 1	...	Stud.j	...	Stud. m	
MCQ 1	a/A	...	d/D	...	t/T	$P_{MCQ1} = 1 - (a/A +...+ d/D +...+ t/T)/m$
...	...	...	...	...	...	...
MCQ i	f/F	...	h/H	...	p/P	$P_{MCQi} = 1 - (f/F +...+ h/H +...+ p/P)/m$
...	...	...	...	...	...	...
MCQ n	s/S	...	u/U	...	y/Y	$P_{MCQn} = 1 - (s/S +...+ u/U +...+ y/Y)/m$
Average Mark 0÷10	cAv.M.1	...	cAv.M.j	...	cAv.M.m	
	For each student Current Average Mark = [1 - (Sum of All Individual Ratios)/n] *10					Only if m = n then: Average P * 10 = Average Mark of all students

Legend: cAv.M.1 means current average mark for student 1 and so on for each student up to last one, student m

For the calculation of the current D index of a MCQ, students in Table III have to be sorted first from left to right together with their ratios for all MCQ from the smallest to the biggest Current Average Mark. This will rank students from the weakest, the first in the left, to the best one, the first in the right. Then, according to definition we mentioned in Introduction, D index equals (P index of the best 27% of the students) minus (P index of the 27% of the weakest/worst students).

We illustrated in Figure 3 our concept of enhanced multiple choice question EMCQ.

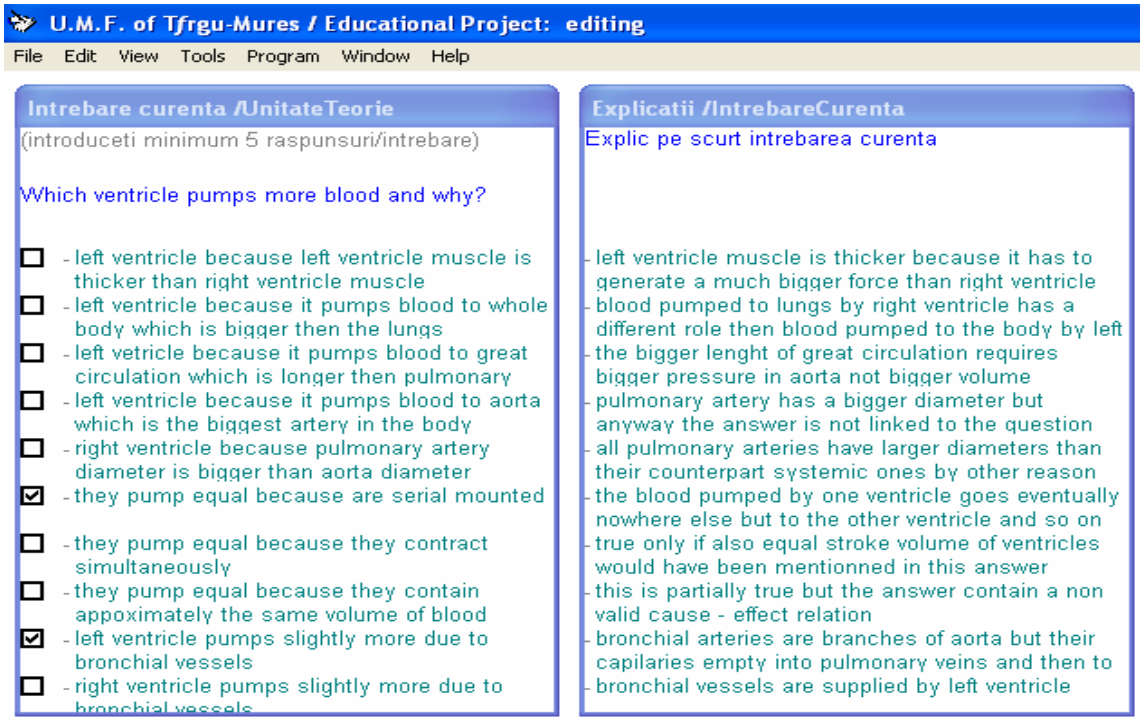

Figure 3. Editing questions in software's Professor Mode. If there are at least 5 answers the question become a ready one to use and software may pick it up to test students

Discussion

The quality of the teacher implication is crucial. Since a lot of students answered wrong to the question in the Figure 3, it is important that any teacher of our physiology department be aware of it and consequently to enable as active only the first 5 'recall' level answers of the MCQ during semester up to a point in time. Our software allows this feature as we explained in methods paragraph. Even the worst student will find out, after trying all the first 5 possible wrong answers, that if it is neither left ventricle nor right ventricle who pumps more than they should pump equal. While they are performing wrong they will also get, at the end of each test, the corresponding explanations for each wrong answer. The clarity of these explanations and how they introduce new physiological terms and mechanisms depends of course on teacher. On the other side if teacher enables as active, by accident, one of the last 'thinking' 5 answers during semester then the first 5 answers, with all the work they carry upon, become useless together with their corresponding explanations.

Students must also be aware of how the concept of EMCQ works and that the teacher will enable answers gradually. Because the 'recall' level is the easiest one in the cognitive domain, even weak students will have enough expectancy of being able to answer correctly the questions and that will motivate them to perform tests. Still, knowing that they will be evaluated on same topics/statements on the final official exam, students will feel that dealing with EMCQ set during semester is of significant value and that will motivate them too in performing the tests. Another source of motivation for the students is that the more times they are performing tests, the more accurately our software can build a top of the most difficult MCQ for each student. Being aware that during the semester they are evaluated only on 'recall' level, students will eventually start the process of thinking of the possible other 5 'understanding' answers they may confront with for a statement they already know.

Someone may claim that students will focus only on MCQ tests and will read no more the written or electronic courses that teachers are providing to them. On the contrary, we believe that

students will use more than now these materials first of all to find out the answers. Then, being aware that during semester the difficulty of an EMCQ will increase, the students will read probably even more. If the students are answering wrong to the questions, it is plausible that they will try to understand the explanations they are given, at the end of the test, for each wrong answer.

Computerized calculation of the current P index and D index of the MCQ set informs teacher about quality of the questions. The teacher may be interested to focus during semester on poor quality questions ie those with a D index less than 0.2 [7,11]. The teacher can also refine the current overall difficulty of the MCQ set to fit PSET index between 0.25 and 0.75 that is neither very easy ($P \geq 0.75$), nor very difficult ($P \leq 0.25$) [7]. Readjusting PSET between 0.4 and 0.6, which it is suggested that spreads out the examinees' scores [12], strongly depends on teacher's competencies. A good teacher can adjust the whole set of EMCQ according to student performances and/or teacher expectancies.

Our software allows the teacher, as it is probably desired towards the middle-end of the semester, to increase the difficulty of an MCQ by enabling as active more than 5 answers. When testing students, the computer will still extract only 5 answers but the number of visual combinations that can result is huge being given by Arrangements of 10 taken by 5, meaning 10*9*8*7*6 that is 30240. This is important because, when testing themselves during semester, our students have the tendency to use visual memory, instead of thinking, in answering questions if they are in static format. This huge amount of visual combinations will obviously decompensate any visual memory if the answers are properly written, for example in length. These aspects should invite students to understand the MCQ and to think of the answers.

The absolute difficulty of MCQs shown in Figure 1, given by the total number of wrong answers from all the students, is useful for the teacher in guiding the education mainly on issues related to those questions. The greater the number of tests performed by students, the more reliable is this P_absolute shown in Figure 1. The fact that our software can generate for each student a similar but individually shaped map helps students to study more on those particular issues from curricula that are the most difficult for each. The possibility of identifying difficult parts of curricula will motivate students to perform more tests during semester.

The correlation coefficient r, between rough D index and rough P index is high enough to appreciate the link as almost strong [9]. The coefficient of determination R^2 shown in Figure 2 is 0.603 and this means that 60.3% of the total variation of items D index is caused by items P index variation or can be explained in relationship with items P index by the polynomial regression equation also shown. The other 39.7% of the total variation of items D index remains unexplained. The variability of D indexes gets greater as the values of P indexes decrease and it is believed that this pattern is not uncommon in biology [13]. The line in Figure 2 is the graphical representation of the regression equation and is useful in determining intervals of P indexes for which D indexes have the greatest values. There are a lot of alternative means of linear regression that can be used if we want to consider variation in both variables: the dependent variable D index and independent variable P index as well [14]. A good MCQ discriminates well students and rise up the spirit of competition which eventually will motivate students to learn and to think. We obtained in our 2012 admittance session biology exam a similar regression line for MCQs previously known to candidates. The more difficult the previously published questions were (low P index), the bigger D index they had, meaning they discriminated students better, even though the students have had previously the possibility to repeatedly test themselves on those MCQs. We believe this is due to the fact that students tried mainly to memorize the answers since it was also known that questions will be exactly the same in the final exam. We already explained how our concept of EMCQ can discourage the students to simply memorize images or texts while answering to questions and how EMCQ concept invites the students to understand and to think as better alternatives.

Conclusions

We developed our existent computer aided concept of teaching and learning medical physiology by adding to our old software new computerized tools that can motivate both teachers and students to work and to think more.

Basic ideas are: to cover well the curricula with the statements of the Enhanced MCQ set, which may have up to ten answers, to enable first the lowest 'recall' cognitive level answers of the EMCQ, to have as much as possible programmed evaluation of students by tests on EMCQ set during semester, to increase later gradually the difficulty of EMCQ set, by enabling the 'understanding' level answers, while monitoring the Current Overall P index and D index of EMCQ set on All Students in a Group, in order to reach teachers expectancies from that group, but maintaining students motivation for self testing also.

The final goal of our computer aided concept is to help teachers to educate well medical students and to give to the teachers tools for proving the quality of the education they performed. We think that the proves consist of the textual content of the EMCQ set the teachers wrote and of students' results at the final evaluation on that set of EMCQ.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Silverthorn DU. Teaching and learning in the interactive classroom. Adv Physiol Educ 2006;30:135-140.
2. Herreid CF. Teaching in The Year 2061. Adv Physiol Educ 2000;24:2-7.
3. Gliga M. Real-time Competition Among Teachers and Students - a Method for Improving Teaching and Assessing Physiology. J Physiol 2005;567P:WA9
4. Petrişor M, Măruşteri M, Ghiga D, Şchiopu A. Online Assessment System. Applied Medical Informatics 2011;28:23-28.
5. IDEAL Consortium. International Database for Enhanced Assessments & Learning [Internet]. [updated unknown; cited 2012 September 16]. Available from: http://www.idealmed.org/homeindex.html
6. The University of North Carolina. Improving Multiple Choice Questions [Internet]. [updated unknown; cited 2012 September 16]. Available from: http://cfe.unc.edu/pdfs/FYC8.pdf
7. Si-Mui Sim, Raja Isaiah Rasiah. Relationship Between Item Difficulty and Discrimination Indices in True/False-Type Multiple Choice Questions of a Para-clinical Multidisciplinary Paper. Ann Acad Med Singapore 2006;35:67-71.
8. Krathwohl, DR. A Revision of Bloom's Taxonomy: An Overview. Theory Pract 2002; 41: 212-264
9. Correlation coefficient and coefficient of determination [Internet]. [updated unknown; cited 2012 September 30]. Available from: http://mathbits.com/MathBits/TIsection/Statistics2/correlation.htm
10. Cognitive processes in motivation [Internet]. [updated unknown; cited 2012 September 16]. Available from: http://www.education.com/reference/article/achievement-motivation/
11. Item Discrimination Indices [Internet]. [updated unknown; cited 2012 September 16]. Available from: http://www.rasch.org/rmt/rmt163a.htm
12. Professional Testing [Internet]. [updated unknown; cited 2012 September 16]. Available from: http://www.proftesting.com/test_topics/steps_9.php
13. Drummond GB, Vowler SL.Categorized or continuous? Strength of an association–and linear regression. Advan in Physiol Edu 2012;36:89-92.
14. Ludbrook J. Linear regression analysis for comparing two measurers or methods of measurement: but which regression? Clin Exp Pharmacol Physiol 2010;37:692-699.