

MULTIPLE-CHOICE TESTS - A TOOL IN ASSESSING KNOWLEDGE

Ana Paula Lopes, Lurdes Babo, José Azevedo, Cristina Torres

Institute of Accounting and Administration of Oporto - ISCAP
Polytechnic Institute of Oporto – IPP
Oporto - Portugal

aplopes@iscap.ipp.pt, lbabo@iscap.ipp.pt, jazevedo@iscap.ipp.pt, ctorres@iscap.ipp.pt

Abstract

The purpose of this paper is to analyse if Multiple-Choice Tests may be considered an interesting alternative for assessing knowledge, particularly in the Mathematics area, as opposed to the traditional methods, such as open questions exams.

In this sense we illustrate some opinions of the researchers in this area.

Often the perception of the people about the construction of this kind of exams is that they are easy to create. But it is not true! Construct well written tests it's a hard work and needs writing ability from the teachers.

Our proposal is analyse the construction difficulties of multiple - choice tests as well some advantages and limitations of this type of tests. We also show the frequent critics and worries, since the beginning of this objective format usage.

Finally in this context some examples of Multiple-Choice Items in the Mathematics area are given, and we illustrate as how we can take advantage and improve this kind of tests.

Keywords - Mathematics, Multiple-Choice Tests, Technology in Education, Teaching Methods.

1 INTRODUCTION

When we hear the word assessment, we know that this is a notable concern of many teachers, especially when we are talking about Testing and Grading. And so many doubts come to our minds. How can we improve the process of gathering, describing or quantify information about the performance of our students? What is the better way to measure the level of achievement or performance of our students?

A specific way to measure performance is testing.

Than some complex questions that we want to be answer emerge like: Are the Multiple-Choice Tests a better alternative for assessing knowledge as opposed to traditional methods, such as open question exams?

In this area the opinions are divergent, but we thought that the Multiple-Choice Tests are also a good way of testing, as any kind of tests they have advantages and limitations. There are particular measurement situations where one Item type is more appropriate than the other.

Some people have many difficulties when they try to do Multiple-Choice Tests and this is one of some reason they are very reluctant in use them. Burton [1] presents another reason:

(...) they believe these items are only good for measuring simple recall of facts. In fact Multiple Choice Items are frequently used to measure lower – level objectives, such as those based on knowledge of terms, facts, methods, and principles. However the real value is their applicability in measuring higher – level objectives, such as those based in comprehension, application, and analysis.

We know they have some limitations, there are things that we can not measuring, certain learning outcomes, such as capable of communication, organization of the information and creativity. It is important to distinguish the objectives that we want assessed. There are objectives which can be appropriately assessed by using Multiple-Choice Tests and others objectives which would be better assessed by other kind of tests (or other means).

However, when we decide to do a Multiple-Choice Test, a lot of doubts arise, such as, how to prepare a good Multiple-Choice Test?

Most of people think that it is easy to construct a Multiple-Choice Test, but it is not, this of course if we want to construct a well written Multiple-Choice Items. Good Multiple-Choice Test Items are generally more complex and time-consuming to create than other types of tests. It requires a certain amount of skill and knowledge. The ability to write Multiple–Choice Items is an important skill for the teacher to develop. This is a situation in which “practice makes perfect” [2]. We have to be sure what we really

want to test, planning carefully the test and the contents that the test will cover before start writing. However, this skill maybe increases through study, training, practice and experience. Even if we don't use this kind of tests as way to evaluate, they could be useful to help students on their individual study.

As it is know the use of the new technologies it is recommended by several organizations (European Parliament and of the Council, The European ODL Liaison Committee, between others), because the information and communication technologies (ICT) offer significant potential for the improvement of education and training [3]. They support learning processes, through enhanced communication, discovery, simulation, exploration and problem solving.

In order to answer these recommendations we have been using Moodle - Modular Object-Oriented Dynamic Learning Environment to the construction of our project started in 2007. Moodle provides teachers with means to create differentiated learning opportunities for students. One part of this project is dedicated to the construction of some kind of tests, in special Multiple-Choice Tests.

2 LEARNING, ASSESSMENT AND EVALUATION

All learning system it is based in teacher-student relationship. In this regard it is necessary to evaluate the knowledge students are acquiring. It is important measurement. To facilitate this evaluation, in most cases, we testing the course that students attended. We think that courses have to integrate the main components, like several procedures, for example, Instruction, Objectives, Assessment and Evaluation.

Different types of assessment can be used. The assessment can be characterized in two different categories: Formative and Summative.

The Formative assessment provides immediate evidence of student learning, purpose to improve quality of student learning and promote modifications in curricular design, and the way we teach. The students receive individual feedback about their strengths and weaknesses.

The Summative assessment occurs most frequently in the final of a course, semester or module. Essentially is most used to make a final decision about the student performance.

Sometimes formative and summative assessments have the same intentions.

The main purposes of the assessment, according to Alison Bone [4], are:

- To grade or rank a student
- To pass or fail a student
- To provide feedback to students
- To provide feedback to lecturers
- To provide feedback to professional bodies
- To contribute to a student profile
- To motivate students
- To motivate lecturers
- To predict success in research and/or professional courses
- To predict success in future employment organization
- To provide a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis for students
- To provide a SWOT analysis for lecturers
- To assist an institution in establishing quality in their provision of courses

To measure the level of achievement or performance of these purposes there are several types of tests, in particular, two general types of test, according Ory [5] are:

- Objective items
- Subjective or essay items

In Objectives items the students have to select one correct answer from de several alternatives, such Multiple-Choice, true-false and matching.

In Subjective or essay Item, also named Constructed-response or open questions, include extended response or essay and restricted-response items, short-answer items or problem solving,

Essay Items are usually easier and faster to build. According to Ory [5] a professional Item writer produces only 9 - 10 good Multiple-Choice Items in a day's time. Also, he cited other authors that find the following:

Both Item types can measure similar content or learning objectives. Research has shown that students respond almost identically to essay and objective test items covering the same content. Studies by Sax & Collet (1968) and Paterson (1926) conducted forty-two years apart reached the same conclusion: "...there seems to be no escape from the conclusions that the two types of exams are measuring identical things." (Paterson, p.246)

Either objective items either essay items are good for measurement the student achievement. But there are some of them most suitable for certain situations assessment of learning situations. In next chapter we will see a hint of better use of different types of valuation for the objectives of learning.

3 BLOOM'S TAXONOMY AND MATH TAXONOMY. TESTING TYPES

As already it was said the assessment and the course objectives are related. The course objectives must be enclosing clear as to what outcomes that students reach in the end of their learning. On the other hand, the instructor must have clearly what intended that students should own at the end of the course. He must write the general and specific objectives and these must have learning levels, from the lower to the higher level of difficulty The Bloom's taxonomy has some suggestions how to construct these objectives and how to develop a hierarchy of learning. The level of knowledge must move from the lower level to the highest.

Bloom's taxonomy still be useful to structure the teaching and learning process.

Bloom's Taxonomy divides educational objectives into three domains: Affective, Psychomotor and Cognitive. Skills in the cognitive domain turn around knowledge, comprehension, and critical thinking of a particular topic. There are six categories in the taxonomy, moving through the lowest order processes to the highest: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation.

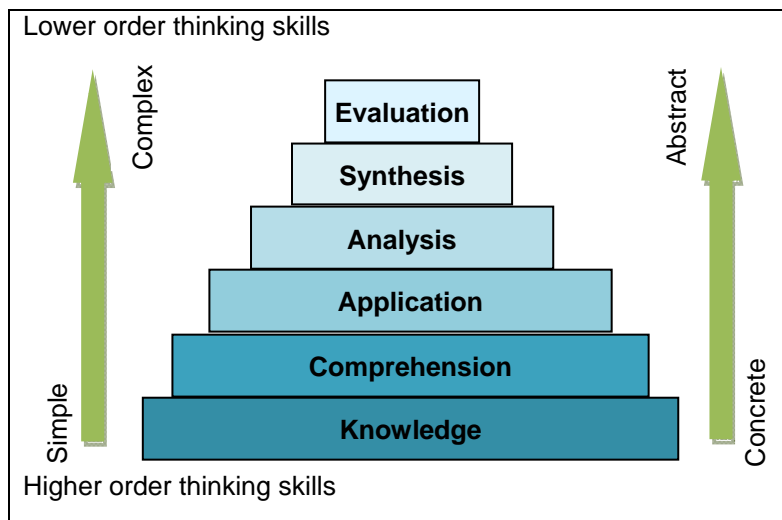


Fig. 1. Bloom's Taxonomy

Based in Bloom's Taxonomy we will propose types of test according to each cognitive level. We can see them in the table 1:

| Cognitive levels | Most Appropriate Test Item |
|------------------------------------|--|
| Knowledge and Comprehension | Multiple-Choice True or False Matching Completion Short answer |
| Application | Multiple-Choice Short answer Problems Essay (Extended-Response) Performance |
| Analysis, Synthesis and Evaluation | Multiple-Choice Short answer Essay |

Table 1. Test Item according to the Bloom Taxonomy

There are several reviews on the use of Multiple-Choice in higher levels. However, how can we see below, it is possible, but more difficult, to construct items in higher level skills. Rodriguez [6], ask: Do Multiple-Choice (MC) Items and construct-response (CR) Items measure the same cognitive behavior? The quick answer is: They do if we write them to do so. Ebel cited in Haladyna [7] says that Alternate-choice Items are not limited to low-level thinking and can be written to measure higher level thinking.

As a way to extending the Bloom’s Taxonomy (Table 1) we suggest a guide teaching and assessing knowledge. For each level have different instructional strategies and testing techniques. To see witch we have to do on each level we suggest to review the following authors: [8], [9], [10], [2], [11] and [12].

Although there are some limitations in this taxonomy, a lot of people use it. There is a specific alternative for Mathematicians. In 1996, a group of Mathematicians, Smith, Wood, Coupland, Stephenson, Crawford, & Ball, from the University of Technology in Sidney have constructed a MATH taxonomy (Mathematical Assessment Task Hierarchy) to solve the problem in structure of assessment tasks.

The MATH taxonomy has three groups’ divides in eight categories (Table 2)

| Group A | Group B | Group C |
|---------------------------|-------------------------------|---|
| Factual knowledge | Information transfer | Justifying and interpreting |
| Comprehension | Application in new situations | Implications, conjectures and comparisons |
| Routine use of procedures | | Evaluation |

Table 2. MATH Taxonomy

These Mathematicians have presented the following conclusion:

“Students enter tertiary institutions with most of their mathematical learning experience in group A tasks, with some experience with group B tasks. Their experience in group C tasks in mathematics is severely limited or non-existent. One of the aims of tertiary education in mathematics should be to develop skills at all three levels.”

Smith et al. [13] recommend the *construct a grid that combines subject topics with the descriptors of the MATH taxonomy*. The grid entries represent a reference to particular questions on the examination paper.

The authors have analyzed many mathematics examination papers that are heavily biased towards group A tasks. A huge number of tests don’t use the group C in higher education. So, this grid is helpful to the professor when he is constructing the test.

We think that we can use both taxonomy’s, in particular the last one can help us in the mathematic items construction.

4 DESCRIPTION OF MULTIPLE–CHOICE TEST ITEMS

A Multiple-Choice test Item consists of two parts:

A problem (stem) – that may be in the form of a question/problem or an incomplete statement, at the beginning of each item.

A list of options (alternatives) – that contains one correct option (the answer) and a number of incorrect options (distractors).

The purpose of the distractors is to appear as tempting solutions to the problem, plausible competitors of the answer for the students that didn’t achieved the objective being measured by the test Item. According to Haladyna [7] a good distractor should be selected by low achievers and ignored by high achievers.

Some authors considerer that Multiple-Choice questions typically have three parts: a stem, the correct answer and the distractors [2], [14].

We can find a variety of Multiple–Choice Items, such as single correct answer, best answer, negative, multiple response and combined response, but we are going to talk about the one we used more – single correct answer.

Below we can see an example of a Multiple-Choice Item in Mathematics area.

Item Stem: There are 10 teachers of Mathematics who work in ISCAP together. Four of these teachers are selected to attend four different conferences in Spain. The first person selected will go to a conference in Madrid, the second will go to Bilbao, the third to Tenerife, and the fourth to Maiorca. How many of such selections are possible?

- Response Alternatives:
- (A) C_4^{10} - Distractor
 - (B) $4!$ - Distractor
 - * (C) P_4^{10} - Answer
 - (D) 4^4 - Distractor

In this case the correct answer is the option who is marked with (*)

Fig. 2. Example of Multiple-Choice Test Item

4.1 HOW TO WRITE A GOOD MULTIPLE-CHOICE ITEM

The most important when we start to writing the question/problem is to give the answer to:

What am I testing?

Once we decided that, we have the guarantee that it goes to test exactly what we want to test, and only what we want to test.

WRITING STEMS

After the student read the stem, he/she should know exactly what the problem is and what he is expected to do to solve it. The question can not be ambiguous, as to be written with clarity, the examinee have to know exactly what is being asked.

The student should not have to read alternatives to understand the question or intent the incomplete statement [2].

What is being assessed can not be the student's ability to infer a description of the problem, but its ability to answer the objective of the Item.

Most the literature shows that we have to be careful with some aspects, such as:

- Identify the one point to be test by that Item
- Include the central idea and most of the phrasing in the stem
- Avoid irrelevant clues, to the correct option
- Eliminate excessive verbiage or irrelevant information
- Restrict the use of negatives in the stem, when used, underline and/or capitalize the negative word

WRITING OPTIONS

Once we have the question, it seems that our task is easier forward, but it is not. Create good alternatives/options it's a hard work.

Downing [15], suggest that

the traditional of using four or five options for Multiple-Choice Items is strong, despite the research evidence suggesting that it is nearly impossible to create selected-response test items with more than about three functional options. (...) This of course makes sense if the test is reasonably well constructed and has a sufficient number of total items which are appropriately targeted in difficulty to the examinees' ability.

The literature shows that to create good alternatives/options we have to be careful with some aspects, for instance:

- They have to be mutually exclusive
- They have to be as homogeneous as possible
- The grammar of each alternative as to be consistent with stem
- Make sure there is only one correct or best response to the Item
- Use plausible distractors
- Incorporate common errors of students in the distractors

- Keep the alternatives parallel in form
- Keep the alternatives similar in length
- Avoid the alternatives “all off the above” and “none of the above” (in general)
- Avoid specific determinates, such as “never” and “always”
- Present the answer in each of the alternative positions approximately as equal number of times, in a random order
- Use letters in front of options rather than numbers
- Distractors that are not chosen by any examinees should be replaced
- Avoid the use of humour when developing options

When all the items are done, the teacher should ask a colleague with expertise in the content area, to review all the items, to make suggestions, for all the problems that could be possible appear, such as ambiguities, redundancies, etc.

4.2 ADVANTAGES AND LIMITATIONS OF MULTIPLE–CHOICE ITEMS

The decision to use Multiple-Choice tests or include Multiple-Choice Items in a test should be based on what the purpose of the test is and the uses that will be made of its results.

We can not forget that there are objectives which can be appropriately assessed by using Multiple-Choice Test Items and others which would be better assessed by some other kind of test Item.

ADVANTAGES

Numerous studies [1], [2], [14], indicate that Multiple–Choice Test Items can be used in many different subject-matter areas, and can be used to measure a great variety of educational objectives. They are adaptable to various levels of learning outcomes, from simple recall of knowledge to more complex levels.

The Multiple–Choice Items are very useful to assessment in large classes. It is also helpful if we intend to implement a system of continuous evaluation based on Multiple-Choice tests performed on the computer. Various works of evaluation can be implemented, automatically corrected and the results exported to Excel.

If we have tools to help our construction of items, like Moodle that provides teachers with a lot of flexibility when creating this common question type, it can be a good way to motivate and help students increasing their independent learning skills. Specially if we give, in online formative test, feedback for any incorrect answer, the students can improve their performance.

Of course there are more advantages that are obvious, like the time of correction, scoring efficiency, accuracy, objectivity and can cover a lot of material very efficiently. Scores are less influenced by guess than true-false items.

The Multiple–Choice Items provides the most useful format if we want to compare the performance from class to class and year to year, always in the same way making use of the objectivity in correction.

LIMITATIONS

As we have been told, it is very difficult to construct well written Multiple-Choice Items.

Multiple-Choice Items needs writing ability from the teachers and reading ability from student's [5]. These tests are difficult to construct particularly at the higher cognitive levels. In general, essay items take less time to write than Multiple-Choice Items, but they are more difficult and time-consuming to score [2].

Like in other kind of tests there are limitations that we have to be aware, Multiple–Choice Test Items can not measuring certain learning outcomes, such as capable of communication and articulate explanations, organization of the information, and creativity - the capable of produce original ideas. Such learning outcomes are better measured by short answer or essay questions, or by performance tests [1].

Sometimes is very difficult to find good distractors.

The students, in Mathematics, become less careful to write the symbolic language.

CRITICS TO THE MULTIPLE–CHOICE TESTS

When we are talking about evaluation with Multiple–Choice Tests, the opinions are divergent. One of the frequent criticism and worried since the beginning of objective format usage, is that students

scores will not fairly represent true achievement unless the scores are transformed in some way to reduce the adverse effects of guessing [16].

In the next table, we can see the faster decreasing of the probability of answer by guessing with the increase of the number of items. For example, in a three Multiple-Choice Items where each has four options, the probability of a student answer by guessing to three items is only in 1.6% of the cases this can happening.

So we can say that from a point of view purely statistical, random guessing alone is extremely unlikely to produce a high test score, obtained a perfect score that can be comparable to the probability of winning the lottery.

| | | Number of Items | | | |
|---------------------------|---|-----------------|-------|-------|-------|
| | | 1 | 2 | 3 | 4 |
| Number of correct answers | 1 | 25% | 37.5% | 42.2% | 42.2% |
| | 2 | | 6.3% | 14.1% | 22.1% |
| | 3 | | | 1.6% | 4.7% |
| | 4 | | | | 0.4% |

Table 3. Percentage of guessing correct answers

Many teachers believe that this kind of questions can measure only memory, and does not give students the necessary freedom of response to measure more complex intellectual abilities.

Opponents of objective testing point out that the essay testing format is a more accurate measure of a student's ability to apply knowledge because it requires students to construct their own response, rather than to simply respond to a proposed answer [10].

However, Multiple-Choice Items written at application and analysis levels require use of concepts and theories and analytical thinking to make selection from the available options. *For items at those levels test-takers need to compare options and make judgment about the correct or best response [2].*

They can be designed so that students have to use critical thinking skills to make the subtle distinctions that are necessary to reason out the correct answer [10]

The zero-tolerance is another of the criticisms that are aimed, and also that in real life, in actual mathematical research, the problems do not usually come with a list of alternatives

5 SOME EXAMPLES

We are going to present some weak and better examples of a few questions in mathematics area. In the following examples the * indicates the correct answer.

Weak example

Calculate the indefinite integral $\int \ln(3x) dx$

(A) $\frac{x}{3}(\ln(3x) - 1) + C, C \in \mathbb{R}$

* (B) $x(\ln(3x) - 1) + C, C \in \mathbb{R}$

(C) $3x(\ln(3x) - 1) + C, C \in \mathbb{R}$

(D) $\frac{1}{x} + C, C \in \mathbb{R}$

Fig. 3. Weak example of MC with no homogeneous alternatives

In the example above the stem follow all the rules that we specified, but when we look to the alternatives we can see that they aren't homogeneous. The student will be inclined to discard the alternative (D).

Next we have a better example.

Better example:

Calculate the indefinite integral $\int \ln(3x) dx$

(A) $\frac{x}{3}(\ln(3x)-1) + C, C \in \mathbb{R}$

* (B) $x(\ln(3x)-1) + C, C \in \mathbb{R}$

(C) $3x(\ln(3x)-1) + C, C \in \mathbb{R}$

(D) $\ln(3x)-1 + C, C \in \mathbb{R}$

Fig. 4. A suggestion of correction for the example of MC with no homogeneous alternatives

Then we have an example where the words “none of the above” were not avoided in constructions of alternatives.

Weak example:

In a certain line of the Pascal Triangle the sum of the last two numbers is 13. What is the sum of the first three numbers of that line

(A) 68

(B) 66

* (C) 79

(D) None of the above

Fig. 5. Weak example of MC with “none of the above”

Imagine that the student has done the calculus and none of results was his result. The question that arises is: he must repeat the calculus because perhaps there’s a little mistake with the calculus or he are going to the option none of the above.

Better example:

In a certain line of the Pascal Triangle the sum of the last two numbers is 13. What is the sum of the first three numbers of that line

(A) 68

(B) 66

* (C) 79

(D) 55

Fig. 6. A suggestion of correction for the example of MC with “none of the above”

The stem should be brief, including only the necessary information. The next example, it’s an example where the stem includes irrelevant information. If the student has to find $\frac{d^2 f}{dy^2}$ and if he got $\frac{df}{dy}$, there is no necessity to know the function $f(x, y)$.

Weak example:

Considerer the function $f(x, y) = x^2 y - 3y^2 x$. Knowing that $\frac{df}{dy} = 2xy - 3y^2$. Then

$\left(\frac{d^2 f}{dy^2}\right)_{\substack{x=0 \\ y=-1}}$ is

(A) 0

(B) 6

(C) 1

* (D) -2

Fig. 7. Weak example of MC with irrelevant information

So, we can write this question in a better way

Better example:

Knowing that $\frac{df}{dy} = 2xy - 3y^2$. Then $\left(\frac{d^2f}{dy^2}\right)_{\substack{x=0 \\ y=-1}}$ is

(A) 0
(B) 6
(C) 1
* (D) -2

Fig. 8. A suggestion of correction for the example of MC with irrelevant information

6 CONCLUSION

In the literature there are many suggestions how to construct a good Multiple-Choice Tests, as well as many opinions about its efficiency in some cognitive levels.

Our experience in Mathematics area at the higher education shows that is possible to construct Multiple Choice Tests in higher level skills. It is more difficult, but it is possible. Although, how we saw in Bloom's Taxonomy, these items and issues are fewer than those of lower level. It is very important that more than one teacher can take part in the construction of multiple choice tests, so several teachers can be involved in the process.

We use the Multiple Choice Tests in Moodle to make continuous assessment. It is very easy and it is very important to give feedback of assessment to the students. They also have, access to formative Multiple-Choice tests that have in each item constant feedback - small suggestions for a resolution or, in most cases, the complete resolution. There is a big database of questions in Multiple-Choice that allow the students accesses, anywhere, to many tests grouped by subjects.

All the above only it is possible with Multiple-Choice Items.

References

- [1] Burton, J. S., [et al] (1991), How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty. Department of Instructional Science. Brigham Young University Testing Services. Web site, 16 August of 2009: <http://testing.byu.edu/info/handbooks/betterItems.pdf>
- [2] Oermann, H. M. & Gaberson, B. K. (2006). Evaluation and Testing in Nursing Education. 2nd ed. Springer. New York - USA.
- [3] Commission of the European Communities. Proposal for a decision of the European Parliament and of the Council adopting a multi-annual programme (2004-2006) for the effective integration of Information and Communication Technologies (ICT) in education and training systems in Europe (eLearning Programme). Web site, 16 August of 2009: http://ec.europa.eu/education/archive/elearning/doc/dec_en.pdf
- [4] Bone, Alison (1999). *Ensuring Successful Assessment*. In Burrige, Roger & Varnava, Tracey (Eds.), Assessment. The National Centre For Legal Education, University of Warwick, Coventry, U.K.
- [5] Ory, C. J. Improving Your Test Questions, Evaluation and Examination Service. University of Iowa. Web site, 16 August of 2009: http://www.uiowa.edu/~examserv/Level_2/resources/Technical%20Bulletins/Tech%20Bulletin%2027.pdf
- [6] Rodriguez, M. C. (2002) *Choosing an Item Format in Large-Scale Assessment Program For All Students: Validity, Technical Adequacy, and Implementation*. Edited by Haladyna, M. T. & Tindal, Gerald. Mahwah, NJ: Lawrence Erlbaum Associates.
- [7] Haladyna, T. M. (1999). *Developing and validating multiple-choice test Items*, 2ns ed. Mahwah, NJ: Lawrence Erlbaum Associates

- [8] Computing Technology for Math Excellence. Teaching and Math Methodology Instruction. Web site, 16 August of 2009: http://www.ct4me.net/math_methodology.htm
- [9] Joining Educational Mathematics. *Bloom's taxonomy of educational objectives*. Web site, 22 August of 2009: <http://www.jem-thematic.net/node/346>
- [10] McDonald, M. E. (2002). *Systematic Assessment of Learning Outcomes: developing multiple-choice exams*. Jones Bartlett Massachusetts - USA.
- [11] Pepin, B. (2008) *Mathematical Tasks in textbooks – Developing an analytical tool based on 'connectivity'*. ICME 11: Mexico. Website, 20 August of 2009: <http://dg.icme11.org/document/get/245>
- [12] Teaching Effectiveness Program. *Bloom's Taxonomy of Cognitive Levels*. Retrieved August, 16, 2009. Web site, 16 August of 2009: <http://tep.uoregon.edu/resources/assessment/multiplechoicequestions/blooms.html>
- [13] Smith, G. H., Wood, L. N., Coupland, M., Stephenson, B., Crawford, K., & Ball, G. (1996). *Constructing mathematical examinations to assess a range of knowledge and skills*. International Journal for Mathematical Education in Science and Technology, 27(1), 65-77.
- [14] Zimmaro, D. M. (2004), *Writing Good Multiple-Choice Exams*. Measurement and Evaluation Center. University of Texas at Austin. Web site, 16 August of 2009: <http://www.utexas.edu/academic/mec/research/pdf/writingmceexamshandout.pdf>
- [15] Downing, M. S. (2006) *Selected-Response Item Formats in Test Development*. In Handbook of test development Edited by Downing, M. S & Haladyna, M. T.I. Mahwah, NJ: Lawrence Erlbaum Associates pp 292
- [16] Downing S. M. (2003), *Medical Education: Guessing on selected-response examinations*. Volume 37 pp 670-671, Chicago, EUA.