US-China Education Review B 1 (2011) 1-11
Earlier title: US-China Education Review, ISSN 1548-6613

Improving Multiple-Choice Questions

Cristina Torres, Ana Paula Lopes, Lurdes Babo, José Azevedo
Institute of Accounting and Administration of Oporto, Oporto, Portugal

A MC (multiple-choice) question can be defined as a question in which students are asked to select one alternative from a given set of alternatives in response to a question stem. The objective of this paper is to analyse if MC questions may be considered as an interesting alternative for assessing knowledge, particularly in the mathematics area, as opposed to the traditional methods, such as open questions. We also present some guidelines with the intent of improving the quality of the MC questions. For that, we illustrate some researchers' opinions in this area. The construction of good MC questions requires special care which needs the writing ability of test-makers and therefore time consuming. We are going to analyse the difficulties in the construction of MC questions as well as some advantages and limitations of this type of questions. We also are showing the frequent critics and worries, since the beginning of this objective format usage. We aim to give some examples of MC items in the mathematics area and illustrate how we can take advantage of and improve this kind of questions in this context.

Keywords: mathematics, MC (multiple-choice) questions, technology in education, teaching methods

Introduction

Assessment is an issue of a notable concern for many teachers, especially when we are talking about testing and grading. Therefore, many doubts and questions come to their minds, such as: How can we improve the process of gathering, describing or quantifying information about the performance of our students? What is the better way to measure the level of the achievement or performance of our students?

A specific way to measure the performance is testing. But meanwhile, a complex question arises: Are the MC (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 MC tests are also a good way of testing. Like any kind of tests, they have advantages and limitations. There are particular measurement situations where one item type is more appropriate than the others.

Some people have many difficulties when they try to make MC items and others are very reluctant in using them. Burton, Sudweeks, Merrill, and Wood (1991) presented a reason for this reluctance: … they believe these items are only good for measuring simple recall of facts. In fact, MC 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 on comprehension, application and analysis.

Cristina Torres, MSc., Institute of Accounting and Administration of Oporto.
Ana Paula Lopes, Ph.D., assistant professor, Department of Mathematics, Institute of Accounting and Administration of Oporto.
Lurdes Babo, MSc., Institute of Accounting and Administration of Oporto.
José Azevedo, MSc., Institute of Accounting and Administration of Oporto.

The MC items have some limitations. There are certain learning outcomes that are difficult to measure with this kind of items, such as the ability to organize the information or the communication and the creativity skills.

It is important to distinguish the objectives that we want to assess. There are objectives which can be appropriately assessed by using MC tests and others which would be better assessed by some other kinds of tests (or other means).

However, how to prepare a good MC test?

The first thought is that it is easy to construct a MC test, but when we think very carefully about constructing a well written MC item some difficulties arise in our mind. Good MC tests are generally more complex and demand a lot of time to create it compared with other types of tests. It requires a certain amount of skills and knowledge. The teacher's ability to write MC items is an important skill to be developed. This is a situation in which "practice makes perfect" (Oermann & Gaberson, 2006). We have to be sure what we really want to assess and plan carefully the contents that the test will cover before the start writing. However, this skill can be increased through study, training, practice and experience.

Even if we do not use this kind of tests as the way to evaluate, they could be useful to help students on their individual studies.

It is known that the use of the new technologies is recommended by several organizations (European Parliament and of the Council, The European ODL (open and distance learning) Liaison Committee, between others), because the ICT (information and communication technologies) offers significant potential for the improvement of education and training (Commission of the European Communities). They support learning processes through enhanced communication, discovery, simulation, exploration and problem-solving.

In order to answer these recommendations, we have been used Moodle (modular object-oriented dynamic learning environment) to the construction of a project since 2007. Moodle provides teachers with tools to create differentiated learning opportunities for students. One part of this project is dedicated to the development of different kinds of tests, in particular MC tests.

Assessment, Evaluation and Measurement

Almost the learning system is based on teacher-student relationship. In this regard, it is necessary to evaluate the knowledge that students are acquiring. It is a very important measurement. 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 assessments can be characterized in two different categories: formative and summative.

The formative assessment provides immediate evidence of students learning, whose the purpose is to improve the quality of students learning and promote modifications in curricular design and the way how we teach. The students receive individual feedbacks about their strengths and weaknesses.

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

Sometimes, the formative and summative assessments have the same intentions.

The main purposes of the assessments, according to Alison and Bone (1999), are: (1) To grade or rank a student; (2) To pass or fail a student; (3) To provide feedbacks to students; (4) To provide feedbacks to lecturers; (5) To provide feedbacks to professional bodies; (6) To contribute to a student profile; (7) To motivate students; (8) To motivate lecturers; (9) To predict success in research and/or professional courses; (10) To predict success in future employment organization; (11) To provide a SWOT (strengths, weaknesses, opportunities and threats) analysis for students; (12) To provide a SWOT analysis for lecturers; and (13) 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 tests, according to Ory (n. d.) are: (1) objective items; and (2) subjective or essay items.

In objectives items, the students have to select one correct answer from several alternatives, such as MC, true-false and matching.

In subjective or essay items, also named constructed-response or open questions, it includes extended response or essay and restricted-response items, short-answer items or problem-solving.

Essay items are usually easier and faster to build whereas objectives items are more time consuming. According to Ory (n. d.), in a professional item writer produces only nine to ten good MC items in a day.

Either objective items or essay items are good for measuring the student achievement. But some of them are most suitable for certain situations, such as assessment of learning situations.

Ory (n. d.) cited other authors that find the following:

> Both item types can measure similar contents or learning objectives. Research has shown that students respond almost identically to essay and objective test items covering the same contents. Studies by Sax and 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)

Bloom's Taxonomy and Mathematical Assessment Task Hierarchy Taxonomies

There is more than one type of learning. As it is said, there is a direct relation between the assessment and the course objectives. The course objectives must be outlined clearly to achieve effective teaching and improve student learning outcomes. If objectives are clearly and well defined, the teacher will have an effective means of evaluating what their students have learned. He/she must write the general and specific objectives which must have learning levels, from the lower to the higher level of difficulty.

The Bloom's taxonomy suggests how to construct these objectives and develop a hierarchy of learning. The level of knowledge must moves from the lower level to the highest.

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

Bloom's taxonomy divides educational objectives into three domains: (1) Affective—growth in feelings or emotional areas (attitude); (2) Psychomotor—manual or physical skills (skills); and (3) Cognitive—mental skills (knowledge).

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 (see Figure 1):

(1) Knowledge—realize students' facility to use memorization and recall some facts;

(2) Comprehension—show students' ability to read course contents and put others' ideas into their own words;

(3) Application—students catch new concepts and apply them to another situation;

(4) Analysis—students get new information and break it down into different parts;

(5) Synthesis—students have the ability to get diverse pieces of information and construct a whole;

(6) Evaluation—involves students' capacity to look at someones' ideas and recognize the importance of the work and the value of the conclusions.

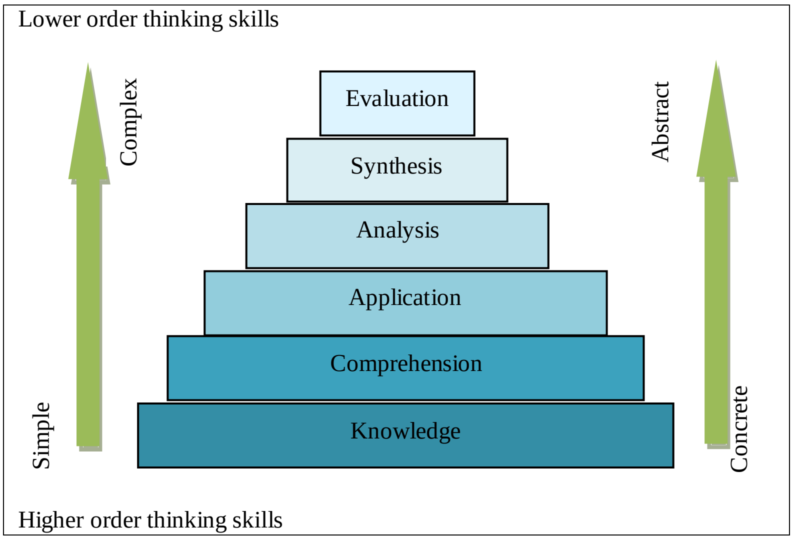

Figure 1. Bloom's taxonomy.

Based on Bloom's taxonomy, we will propose different types of tests according to each cognitive level. We can observe them in Table 1.

Table 1

Test Item According to the Bloom's Taxonomy

Cognitive levels	Most appropriate test item
Knowledge and comprehension	MC True or false Matching Completion Short answer
Application	MC Short answer Problems Essay (Extended-response) Performance
Analysis, synthesis and evaluation	MC Short answer Essay

There are various opinions about the use of MC at higher levels. However, how can we see below, it is possible, but more difficult to construct items in higher-level skills. Rodriguez (2002) wondered if the MC items and CR (construct-response) items measured the same cognitive behaviour. The answer is affirmative "They do if we write them to do so".

Ebel cited in Haladyna (1999) said that alternate-choice items were not limited to low-level thinking and could be written to measure higher level thinking.

As a way to extending the Bloom's taxonomy (see Table 1), we suggest a guide teaching and assessing knowledge. Each level has different instructional strategies and testing techniques (Computing Technology for

Math Excellence, n. d.; Joining Educational Mathematics, n. d.; McDonald, 2002; Oermann & Gaberson, 2006; Pepin, 2008; Teaching Effectiveness Program, n. d.).

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

The MATH taxonomy has three groups divided into eight categories (see Table 2).

Table 2

MATH Taxonomy

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

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 in 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 these three levels. (p. 71)

Smith et al. (1996) recommend to 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's tasks. A huge number of tests do not use the group C in higher education. So, this grid is helpful for the professor when he is constructing the test.

We think that we can use both of the taxonomies, in particular, the last one can help us in the mathematic items construction.

Description of MC Question

A MC question consists of two parts: (1) The problem or stem that may be in the form of a question/problem or an incomplete statement, at the beginning of each item, and followed by the options; and (2) A set of options or alternatives that contains one correct option, the answer or key and several incorrect options, called distractors.

The purpose of the distractors is to appear as tempting solutions to the problem, plausible competitors of the answer for the students that do not achieved the objective measured by the test item. According to Haladyna (1999), a good distractor should be selected by low achievers and ignored by high achievers.

Some authors considered that MC questions typically have three parts: the stem, the correct answer and the distractors (Oermann & Gaberson, 2006; Zimmaro, 2004).

We can find a variety of MC 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.

Figure 2 illustrates an example of a MC item in mathematics area.

Item stem: A teacher plans to construct a 25-question exam from previous exam, in which there are 50 true-false questions and 80 multiple-choice questions. Ignoring the order of the questions, how many exams can be constructed?

Response alternatives:

(A)	P_{25}^{130}	- Distractor
(B)	25!	- Distractor
(C)*	C_{25}^{130}	- Answer
(D)	25^{25}	- Distractor

Figure 2. Example of MC test item. Note. In this case, the correct answer is the option which is marked with *.

How to Write a Good MC Question

We must keep in mind that one of the most important things when we start to write the question/problem/stem is to give the answer to: What am I testing?

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

Suggestions for writing stems. When we are writing stems we must be sure that the question posed is clear but not ambiguous, the examinee have to know exactly what is being asked.

The student should know exactly what the problem was after reading the item and should not have to read alternatives to understand the question or intent of the incomplete statement (Oermann & Gaberson, 2006).

What is being assessed can not be the students' ability to infer a description of the problem, but the ability to answer the objective of the item.

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

(1) Begin writing items well ahead of the time when they will be used allowing time for revision;

(2) Identify the one point to be tested by that item;

(3) Include the central idea and most of the phrasing in the stem;

(4) Avoid irrelevant clues to the correct option;

(5) Eliminate excessive verbiage or irrelevant information;

(6) Restrict the use of negatives in the stem, when used, underline and/or capitalize the negative word;

(7) Questions should follow an easy to difficult progression.

Suggestions for writing options. Once we have the question, it seems that our task is easier forward, but it is not. Creating good alternatives/options is a hard work.

Downing (2006) suggested 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' abilities. (p. 292)

The literature shows that to create good alternatives/options we must pay attention to some aspects, for example:

(1) Be brief;

(2) Avoid overlapping alternatives;

(3) Be mutually exclusive;

(4) Avoid excessive use of negatives or double negatives;

(5) Be more homogeneous if possible;

(6) The grammar of each alternative to be consistent with stem;

(7) Make sure there is only one correct or best response to the item;

(8) Use plausible distractors;

(9) Incorporate common errors of students in the distractors;

(10) Keep the alternatives parallel in form;

(11) Keep the alternatives similar in length;

(12) Avoid the alternatives "all of the above" and "none of the above" (in general);

(13) Avoid specific determinates, such as "never" and "always";

(14) Present the answer in each of the alternative positions approximately as equal number of times, in a random order;

(15) Use letters in front of options rather than numbers;

(16) Distractors that are not chosen by any examinees should be replaced;

(17) Avoid the use of humour when developing options;

(18) The position of the correct answer should vary randomly from item to item;

(19) Develop a lot of questions to generate several tests and distribute them randomly to avoid cheating by students.

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

Advantages and Limitations of MC Questions

MC questions have advantages and limitations just as any other kind of test item.

The decision to use MC tests or include MC items in a test should be based on what the purpose of the test is and the use that will be made of its results.

We can not forget that there are objectives which can be appropriately assessed by using MC test items and others which would be better assessed by some other kinds of test items.

Some advantages. Several studies (Burton et al., 1991; Oermann & Gaberson, 2006; Zimmaro, 2004) indicated that MC test items could be used in many different subject-matter areas, and could 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 difficult levels.

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

If we have tools to help our construction of items, such as 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 increase their independent learning skills. Especially, if we give feedback for any incorrect answer in online formative test, the students can improve their performances.

There are some advantages that are obvious, like the time of correction, scoring efficiency, accuracy, objectivity and covering a lot of material very efficiently. Scores are less influenced by guess the true-false items.

The MC items provide the most useful format if we want to compare the performances from class to class and year to year, always in the same way making use of the objectivity in correction.

Some limitations. As we have been told, it is not easy to develop well written MC items. MC items need writing ability of the teachers and reading ability of the students (Ory, n. d., p. 7). These tests are difficult to construct particularly at the higher cognitive levels. In general, essay items take less time to write than MC items, but they are more difficult and time-consuming to be scored (Oermann & Gaberson, 2006, p. 98).

Just as any other type of tests, there are limitations that we have to be aware. MC test items can not measure certain learning outcomes, such as the ability to communicate and articulate explanations, organization of the information, and creativity (the capacity of produce original ideas). Such learning outcomes were better measured by short answer or essay questions, or by performance tests (Burton et al., 1991).

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

Another limitation is that students cannot justify their choices.

Sometimes, it is very difficult to develop good distractors.

MC questions—critical thinking. MC tests may be considered as an interesting alternative for assessing, but there are some divergent opinions. One of the frequent criticism and worry since the beginning of objective format usage, is that students' scores not fairly represent true achievement unless the scores are transformed in some way to reduce the adverse effects of guessing (Downing, 2003).

In Table 3, we can see the faster decreasing of the probability of answering by guessing with the increase of the number of items. For example, in three MC items where each has four options, the probability of a student answer by guessing to three items is only 1.6% of the cases that can happen.

So we can say that from a point of view purely statistical, random guessing alone is extremely unlikely to produce a high test score.

Table 3

Percentage of Guessing Correct Answers

		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%

Many teachers believe that this kind of questions can measure only the memory, but 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 measurement of a student's ability to apply knowledge because it requires students to construct their own responses, rather than to simply respond to a proposed answer. (McDonald, 2002, p. 84)

However, MC items written at application and analysis levels require the 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 (Oermann & Gaberson, 2006).

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 (McDonald, 2002).

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

Another issue is that MC tests are more susceptible to certain types of cheating than other type of tests, since the correct answer is easy to copy, even by students who do not really understand the subject.

Some Examples

We are going to present some poor and good examples of multiple-choice questions in mathematics area.

In the following examples, the * indicates the correct answer.

The poor example is shown in Figure 3,

Calculate the indefinite integral $\int \ln(4x)\,dx$

(A) $\frac{x}{4}\left(\ln(4x)-1\right)+C,\ C\in\mathrm{R}$

(B)* $x\left(\ln(4x)-1\right)+C,\ C\in\mathrm{R}$

(C) $4x\left(\ln(4x)-1\right)+C,\ C\in\mathrm{R}$

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

Figure 3. The poor example of MC with no homogeneous alternatives. Note. In this case, the correct answer is the option which is marked with *.

As it is seen in Figure 3, the stem follows all the rules that we have specified, but when we look at the alternatives we can see that they are not homogeneous. The student will be inclined to discard the alternative (D).

Next, we have a better example (see Figure 4).

The good example is,

Calculate the indefinite integral $\int \ln(4x)\,dx$

(A) $\frac{x}{4}\left(\ln(4x)-1\right)+C,\ C\in\mathrm{R}$

(B)* $x\left(\ln(4x)-1\right)+C,\ C\in\mathrm{R}$

(C) $4x\left(\ln(4x)-1\right)+C,\ C\in\mathrm{R}$

(D) $\ln(4x)-1+C,\ C\in\mathrm{R}$

Figure 4. A suggestion of correction for the example of MC with no homogeneous alternatives. Note. In this case, the correct answer is the option which is marked with *.

In Figure 5, we have an example where the words "none of the above" were not avoided in constructions of alternatives.

The weak example is,

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

Figure 5. Weak example of MC with "none of the above". Note. In this case, the correct answer is the option which is marked with *.

Imagine that the student has done the calculus and none of the results was his/her result. The question that arises is: He must repeat the calculus because there may be a few mistakes with the calculus or he/she is going to choose the option none of the above.

The good example (see Figure 6) is,

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

Figure 6. A suggestion of correction for the example of MC with "none of the above". Note. In this case, the correct answer is the option which is marked with *.

The stem should be brief, including only the necessary information. The next example (see Figure 7), it is an example where the stem includes irrelevant information. If the student has to find $\frac{d^2 f}{dy^2}$ and if he/she got $\frac{df}{dy}$, there is no necessity to know the function $f(x, y)$.

The poor example is,

Considerer the function $f(x,y) = 2x^3y - 4y^2x$. Knowing that $\frac{df}{dx} = 6x^2y - 4y^2$. Then $\left(\frac{d^2 f}{dx^2}\right)_{\substack{x=0\\y=-1}}$ is
(A) 12
(B) 6
(C) 1
(D)* 0

Figure 7. Poor example of MC with irrelevant information. Note. In this case, the correct answer is the option which is marked with *.

So, we can write this question in a better way (see Figure 8).

The good example is,

Knowing that $\frac{df}{dx} = 6x^2y - 4y^2$. Then $\left(\frac{d^2 f}{dx^2}\right)_{\substack{x=0\\y=-1}}$ is
(A) 12
(B) 6
(C) 1
(D)* 0

Figure 8. A suggestion of correction for the example of MC with irrelevant information. Note. In this case, the correct answer is the option which is marked with *.

Conclusion

In the literature, we can find many suggestions about how to construct/develop a good MC question, as well as many opinions about its efficiency at some cognitive levels.

Our experience, in mathematics area, at the higher education shows that it is possible to construct MC questions in higher-level skills which is more difficult, but possible. Although, we saw that in Bloom's taxonomy, these items and issues were fewer than those at lower level. It is very important that more than one teacher can take part in the construction of MC tests, so that several teachers can be involved into this process.

We use the MC tests in Moodle to make continuous assessment. It is very easy and important to give feedbacks of the assessment to students. They also have, access to formative MC tests that have in each item constant feedback—small suggestions for a resolution or, in most cases, the complete resolution. There is a huge database of MC questions in Moodle that allows us to make random tests about different topics. The use of Moodle allows students to access at anywhere and anytime to different randomly tests, grouped by themes. We think that MC tests can play a very useful supporting role in self-evaluation for almost all math courses, in particular with regards to foundational materials such as definitions or basic rules of calculation.

References

Bone, A. (1999). *Ensuring successful assessment*. In R. Burridge, & T. Varnava. (Eds.), *Assessment*. Coventry: The National Centre for Legal Education, University of Warwick.

Burton, J. S., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to prepare better multiple-choice test items: Guidelines for university faculty*. Department of Instructional Science, Brigham Young University Testing Services. Retrieved August 16, 2009, from http://testing.byu.edu/info/handbooks/betterItems.pdf

Commission of the European Communities. (n. d.). *Proposal for a decision of the European parliament and of the council adopting a multi-annual programme (2004-2006) for the effective integration of ICT (information and communication technologies) in education and training systems in Europe (e-learning programme).* Retrieved August 16, 2009, from http://ec.europa.eu/education/archive/elearning/doc/dec_en.pdf

Computing Technology for Math Excellence. (n. d.). *Teaching and math methodology instruction*. Retrieved August 16, 2009, from http://www.ct4me.net/math_methodology.htm

Downing, S. M. (2003). *Medical education: Guessing on selected-response examinations* (Vol. 37, pp. 670-671). Chicago, EUA.

Downning, S. M. (2006). *Selected-response item formats in test development*. In M. S. Downning, & M. T. Haladyna. (Eds.), *Handbook of test development*. Mahwah, N. J.: Lawrence Erlbaum Associates, p. 292.

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items* (2nd ed.). Mahwah, N. J.: Lawrence Erlbaum Associates.

Joining Educational Mathematics. (n. d.). *Bloom's taxonomy of educational objectives*. Retrieved August 22, 2009, from http://www.jem-thematic.net/node/346

McDonald, M. E. (2002). *Systematic assessment of learning qutcomes: Developing multiple-choice exams*. USA: Jones Bartlett Massachusetts.

Oermann, H. M., & Gaberson, B. K. (2006). *Evaluation and testing in nursing education* (2nd ed.). Springer, N. Y.

Ory, C. J. (n. d.). *Improving your test questions, evaluation and examination service*. University of Iowa. Retrieved August, 16, 2009, from http://www.uiowa.edu/~examserv/Level_2/resources/Technical%20Bulletins/Tech%20Bulletin%2027.pdf

Pepin, B. (2008). *Mathematical tasks in textbooks—Developing an analytical tool based on "connectivity"*. ICME 11: Mexico. Retrieved August 20, 2009, from http://dg.icme11.org/document/get/245

Rodriguez, M. C. (2002). *Choosing an item format in large-scale assessment program for all students: Validity, technical sdequacy and implementation* (M. T. Haladyna, & G. Tindal, (Eds.)). Mahwah, N. J.: Lawrence Erlbaum Associates.

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.

Teaching Effectiveness Program. (n. d.). *Bloom's taxonomy of cognitive levels*. Retrieved August 16, 2009, from http://tep.uoregon.edu/resources/assessment/multiplechoicequestions/blooms.html

Zimmaro, D. M. (2004). *Writing good multiple-choice exams*. Measurement and Evaluation Center. University of Texas. Retrieved August 16, 2009, from http://www.utexas.edu/academic/mec/research/pdf/writingmcexamshandout.pdf