DESIGNING MULTIPLE CHOICE TESTS FOR BUSINESS ENGINEERING STUDENTS AND TRAINEE TRANSLATORS

Associate Professor Suzana Carmen Cismas
Department of Modern Languages
The University of Agronomic Sciences and Veterinary Medicine, Bucharest, Romania

Abstract
Multiple-choice tests are a rapid and clear framework to assess all learners' grasp of a topic. It is a method of retrieval and practice, hence it assists in improving candidates' retention and recall abilities. Subsequent to multiple-choice tests students may get false information. Research has consistently shown that giving feed-back eliminates misunderstandings and improves performance levels. An increasingly popular add-on to such exams is enforced self-confidence. Research shows that using such exam features makes test-takers like assessment and develop effective study skills. They prove of particular use in the acquisition of business engineering concepts in English, both in foreign language learning and in CLIL taught at academic levels.

Key words: multiple choice test design, memory, higher order thinking skills, business engineering English

To outline and clarify the background, multiple choice items are only one section within the framework of test questions available in task banks for teachers' ready access, with advantages and disadvantages as summarised next:

- MULTIPLE-CHOICE QUESTIONS
  Advantages: quick and easy to score, by hand or electronically; can be written so that they test a wide range of higher-order thinking skills; covers large content areas in a 50-minute exam.
  Disadvantages: tests literacy skills (by reading a question carefully, the answer can come easy with little study); gives unprepared students chances to guess&score for unfair credit; exposes students to misinformation; it influences later thinking on the content; takes time &skill to draft.
- TRUE-FALSE QUESTIONS
  Advantages: quick and easy to score
  Disadvantages: unreliable, favouring guesses; the statement is mostly true, except a small or trivial detail, which makes it all false in the end.
- SHORT-ANSWER QUESTIONS
  Advantages: quick and easy to grade and write
  Disadvantages: favours memorization; students' understanding of content remains superficial.
- ESSAY QUESTIONS
  Advantages: students demonstrate knowledge, skills, and abilities in a variety of ways; develop writing skills, formulate arguments backed by reasoning and evidence
  Disadvantages: need extensive time to grade; allow subjective criteria when assessing answers; requires quick composition without planning & revision time, which yields poor-quality work
- QUESTIONS PROVIDED BY TEST BANKS
  Advantages: saves time&energy for drafting test items; uses terms&methods from the text book.
  Disadvantages: rarely test analysis/synthesis/application/evaluation (85% items assess memory) limits exam scope to text content; leads students to view class material as unimportant.

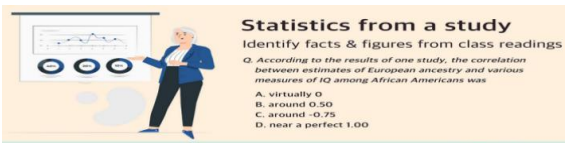
	Benefits	Weaknesses
Yes/No (True/False) questions	• Simple questions • Question and statements covered. • Can be used in different way (positive, negative and combined)	• Question design • Tests with only this type of question are hard
Multiple choice	• Use in different way (one right answer, multiple correct answers) • Most common type of questions	• Require additional design consideration since not all answers can have the same weight • It is not always easy to prepare the answers that are not trivial to dismiss
Matching	• Dynamic questions • Require more than just clicking • Can be used for matching or categorizing	• Not all authoring tool enables the classification (multiple answers to put into the same category) • Not all browsers are capable to handle moving objects
Short answer	• Appropriate in language learning	• Not all authoring tools enable multiple correct answers. • Sometimes correct answer can be lost.
Numerical	• More suitable for numerical answer then text • Type checking	• Sometimes when the result is real number true answers can be lost because of rounding
Essay	• Enables creativity of student	• Cannot be electronically evaluated • Post processing is required • Response is delayed • Due to technical limitation sometimes special characters can corrupt data

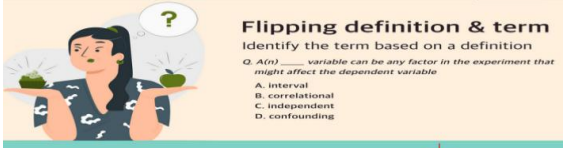
Characteristics	Example
Convergence	Insulin leads to the dephosphorylation and activation of FBPase-2. Which of the following would be a direct effect of this activation? A. **Activation** of fatty acid synthesis. B. **Activation** of glycogen synthesis. C. **Activation** of glycolysis. D. **Activation** of gluconeogenesis. E. Inhibition of gluconeogenesis. Alternatives converge on "activation," suggesting it is part of the correct answer.
Paired answers	Insulin leads to the dephosphorylation and activation of FBPase-2. Which of the following would be a direct effect of this activation? A. Activation of fatty acid synthesis. B. Activation of glycogen synthesis. C. Activation of glycolysis. D. **Activation of gluconeogenesis.** E. **Inhibition of gluconeogenesis.** Paired choices suggest one is correct.
Grammatical cues	In mitochondria that have been treated with the ATP synthase blocker oligomycin, O_2 consumption can be restored by administration of _____. A. Ammonium. B. Greater activity within the rest of the cell. C. Paranitrophenol. Grammar is inconsistent with stem, making this alternative implausible.
Answers are different lengths	When performing a Grignard reaction, why is it important to ensure there is *no* water present? A. Grignard reagents irreversibly react with water, causing reaction yield to be lower. B. Magnesium complexes with water. C. Water may react with the alkyl group. One answer is longer, a cue that this answer is correct.
Question wording cues right answer	The process of an atom gaining or losing an electron to form an **ion** is referred to as _____. A. **Ion**ization. B. Oxidation. C. Reduction. D. Vaporization. The word ion in the stem cues students to the correct alternative.

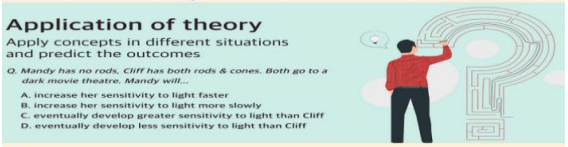
Anatomy of a formative MCQ

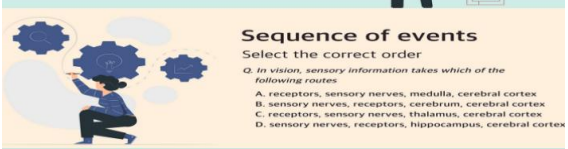
Fig. 1. Profiles of question types with examples and scoring, cf. Ebel&Frisbie, 1986

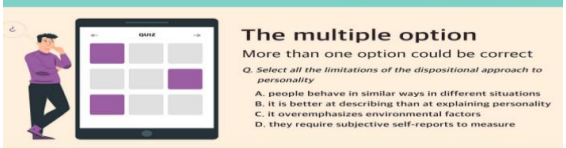
Multiple-choice tests are useful evaluation tools and effective learning strategies, both in the traditional classroom and in CLIL academic environments. Certain benefits of using this type of retrieval practice include: identifying knowledge gaps, making&strengthening connections between concepts/processes, checking for misunderstandings and enforcing personal progress under pressure.

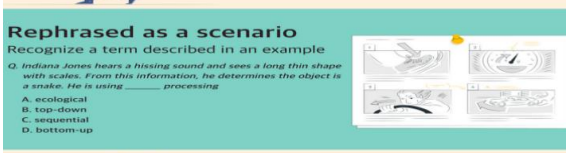












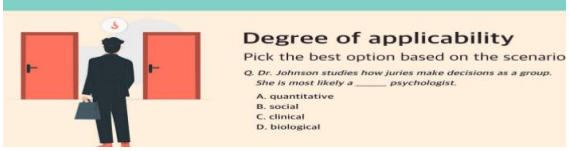


Fig. 2. The 10 types of multiple choice questions cf. Acadly, 2017

Multiple choice tests are efficient in assessing learning outcomes and have several advantages:

- Versatility is an important feature. Multiple choice test items are drafted to assess different levels of learning outcomes, from recall to application, analysis, and evaluation. As students choose from a set of potential answers, however, there are obvious limits on what can be tested with multiple choice items. They are not an effective way to test learners' ability to organize thoughts or articulate explanations or creative ideas.
- Reliability: Reliability is defined as the degree to which a test consistently measures a learning outcome. Multiple choice items are less susceptible to guessing than true/false questions, hence they are more reliable assessment means. Reliability is enhanced when the number of multiple choice items focused on a single learning objective is increased. In addition, the objective scoring associated with multiple choice test items frees them from problems with scorer inconsistency that may distort evaluation in essay questions.
- Validity is the degree to which a test measures the learning outcomes it needs to scale. Since students typically answer a multiple choice item faster than an essay question, tests based on multiple choice items can typically focus on a relatively broad representation of course material, thus increasing assessment validity.

The key to taking advantage of such strengths is formulating clear multiple choice items. It consists of a problem, known as the stem, and a list of suggested solutions, known as alternatives. The latter consist of one correct/best option, which is the answer, and incorrect/inferior alternatives, known as distractors. The stem must be able to stand alone as a short-answer question without alternatives. The stem should either be in a question format or completion format, but the question format is recommended by experts. In the completion stem, the blank to be answered must always be at the end. The stem always includes a verb. Distractors are meant to offer feasible but inaccurate, incomplete, or less accurate options that a student who does not know or has a partial understanding of the material may select. Typically, a question has 3-4 distractors. Research says that more than 4 yield little benefit. Less than 3 distractors will improve the odds of guessing. All distractors should be homogeneous in content, form, and grammatical structure. Each must be unique and plausible. If nobody ever picks a distractor as the answer, it is not a good distractor. One distractor must not be rephrased to create another one. Never use synonyms or homographs to make distractors.

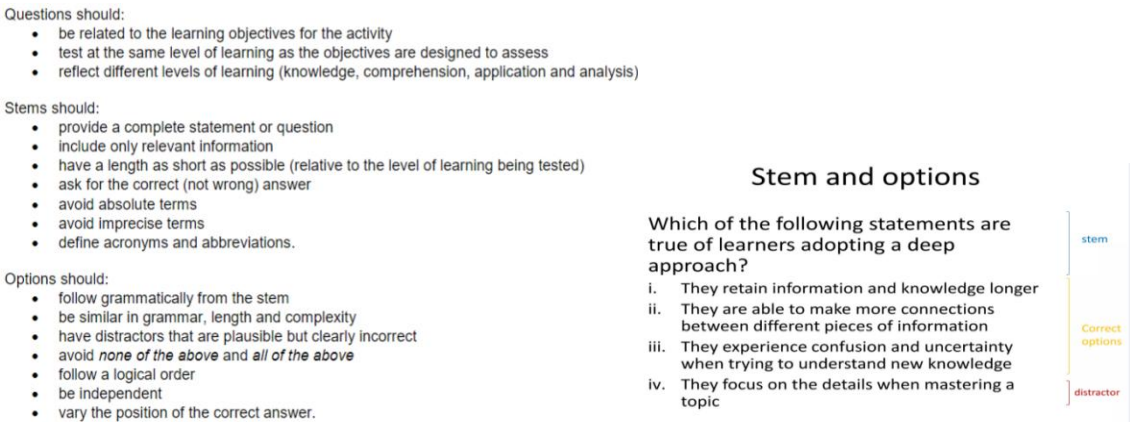


Fig. 3. Guidelines for designing multiple choice questions, with sample item, cf. Aronson, 2002

- Constructing an effective stem:

1. It must be meaningful by itself &present a clear problem enabling focus on the study aim. It serves as a direct test of students' achievement of the learning outcome, not a test for their ability to draw inferences from vague descriptions.

2. The stem should not contain irrelevant material, which can decrease the reliability and the validity of the test scores (Haldyna & Downing 1989).

3. The stem is negatively stated only when significant learning outcomes need it. Students have difficulty understanding negative items (Rodriguez, 1997). If negative phrasing is required, (e.g. identification of dangerous lab/clinical practices), the negative element is emphasized with italics or capitalization.

4. The stem is a question/partial sentence.A question stem is preferred because it allows the student to focus on answering the item rather than holding the partial sentence in working memory and sequentially completing it with each alternative. The cognitive load is bigger if the stem is built with an initial or interior blank, hence it should be avoided.

- Constructing effective alternatives

1. All alternatives must be plausible. Incorrect alternatives serve as distractors, which will be selected by students who did not achieve the learning outcome but ignored by the ones who did that. Implausible alternatives are not good distractors, so they are not to be used; common student errors make the best distractors.

2. Alternatives will be stated clearly &concisely; if excessively wordy, they assess students' reading ability, rather than their attainment of the objective.

3. Alternatives are mutually exclusive. The ones with overlapping content are seen as tricks by test-takers, and erode both trust and respect for the testing process.

4. Alternatives must be homogeneous in content. Those with heterogeneous content could provide cues to student about the correct answer.

5. Alternatives must not contain indications about the correct response. Sophisticated test-takers are alert to inadvertent clues, such as differences in grammar, length, formatting and language choice in the alternatives. Therefore alternatives need: grammar like the stem, parallel forms, similar length, similar language.

6. Alternatives like *all of the above* and *none of the above* should not be used because test-takers who can identify more than one alternative as correct can select the correct answer even if they are unsure about other alternatives.When *none of the above* is used, candidates who can eliminate a single option can thereby eliminate a second one, thus using partial knowledge to get a correct answer.

7. Alternatives should be presented in logical order (alphabetical or numerical) to avoid bias toward certain positions.

8. The number of alternatives varies among items if all options are plausible. The plausible ones are functional distractors, and are chosen by students with imprecise knowledge, but are ignored by those who have achieved the aim. There is little difference in difficulty, discrimination and test score reliability among items with two, three, and four distractors.

- Additional Guidelines

1. Avoid complex multiple choice items, where some/all alternatives consist of different combinations of options. As with *all of the above*, a skilled test-taker could use partial knowledge to achieve a correct answer.

2. Keep the specific content of items independent of one another. Savvy test-takers use data in a question to answer another one, thus reducing test validity.

There are many factors to consider in conceiving a multiple-choice test: how many options should be included; how difficult should the questions be; if/when to give feedback.

The focus is on a stem that presents a problem requiring application of course principles, analysis of a problem, or evaluation of alternatives; these envisage higher-order thinking and thus test students' ability to use it. Designing alternatives requiring a high level of discrimination can also contribute to multiple choice items that test higher-order thinking. Selecting problems needing multi-logical approaches are another avenue to explore: 'thinking that requires knowledge of more than one fact to logically and systematically apply concepts to a problem' (Morrison, 2001).Testing higher-order thinking is paramount in drafting multiple choice tests. In designing such items to test elaborate thinking, draft items requiring elevated cognition levels as seen in Bloom's taxonomy:

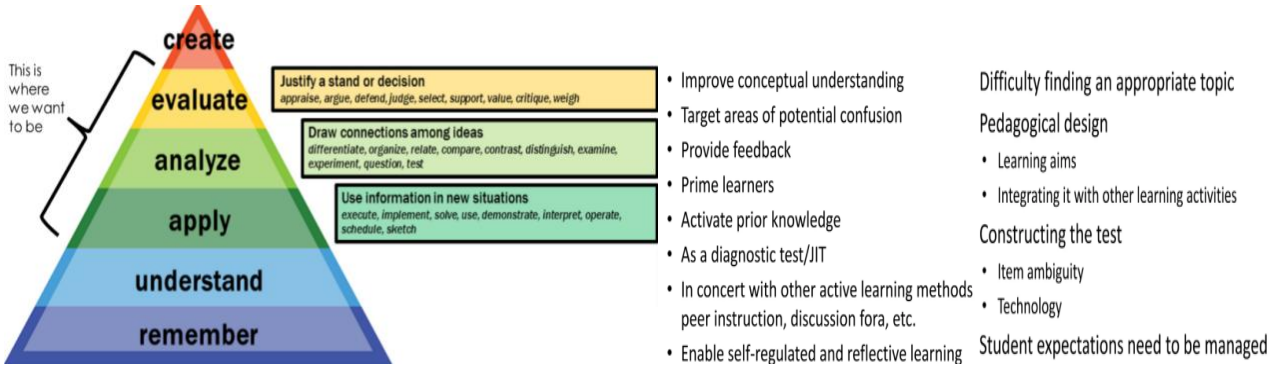


Fig. 4. Bloom's taxonomy, cf. www.simplypsychology.org

Perhaps there is no such thing as the perfect multiple-choice test, but researchers exploring this issue have suggested certain guidelines to consider, illustrated next:

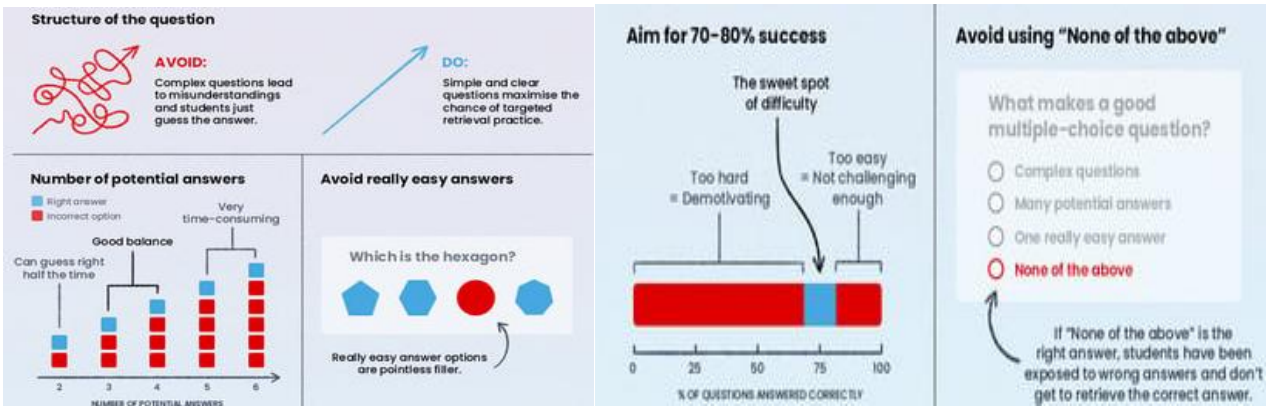


Fig. 5. Designing the perfect multiple choice test, cf. Butler 2019, via www. innerdrive.co.uk

- Avoid complex questions and answer formats. If the question is complicated, it becomes less about how well students can retrieve information, but, more likely, about guessing the answer. Complexity feeds ambiguity.The evaluator does not know if candidates really knew the answer or if they were lucky. The same happens when students are given complex response methods, such as answer-until-correct method (if students get the answer wrong on the first attempt, they can continue until they find the correct answer). However, research shows no benefits to using this method over having a traditional multiple-choice test with immediate feedback.
- Create questions with specific cognitive processes in mind. Each item included should measure a specific part of knowledge or thought process. For example, multiple-choice questions may be requiring students to retrieve a fact, to contrast topics, or to apply a theory to a new situation. Thus testing assists in building learners' schemata, therefore increasing their rate of learning.

- Avoid *none of the above/all of the above* options. It is prudent eliminate such options because they increase exposure to false answers. As research shows, it can result in a negative impact on students' performance since they continue to falsely select 1 incorrect answer. On the other hand, if the *all of the above* option is correct, it is useful because learners are exposed to all the correct answers. However, if this option is not right it results in students believing that multiple false alternatives are true. Hence both of these options might be detrimental to learning.
- It is best to give 3-4 options. Meta-analyses found that having 3options is optimal: one correct answer and two lures is the best balance between quality &efficiency. However, it also depends on the number of plausible incorrect options that can be made.
- Have moderately difficult tests, to challenge, not to discourage students. Researchers suggest that learners must correctly answer 75%, which is higher than the midpoint between what they can get by chance or perfect performance. Challenging learners and showing high expectations significantly improve academic results, because students subsequently raise their effort levels to live up to someone else's high standards. If work is too challenging, it will have a negative impact on their self-confidence. Therefore, finding the right balance is very important.
- Feedback is a key aspect when creating multiple choice items, to combat the false information candidates may get after a test. Research has consistently demonstrated that giving feedback helps eliminate misunderstandings and will improve students' performance.

Advice on designing tasks includes starting with a question and revising it until it can be answered in one word/short phrase. Do not list answer options for that single question, but attach to the exam an alphabetized answers list. Students select options from that list. Some of the ones provided may be used more than once, some may not be used at all, and there are more answers than questions. It is a version of matching; it makes the test challenging&decreases the chance of scoring by guessing.

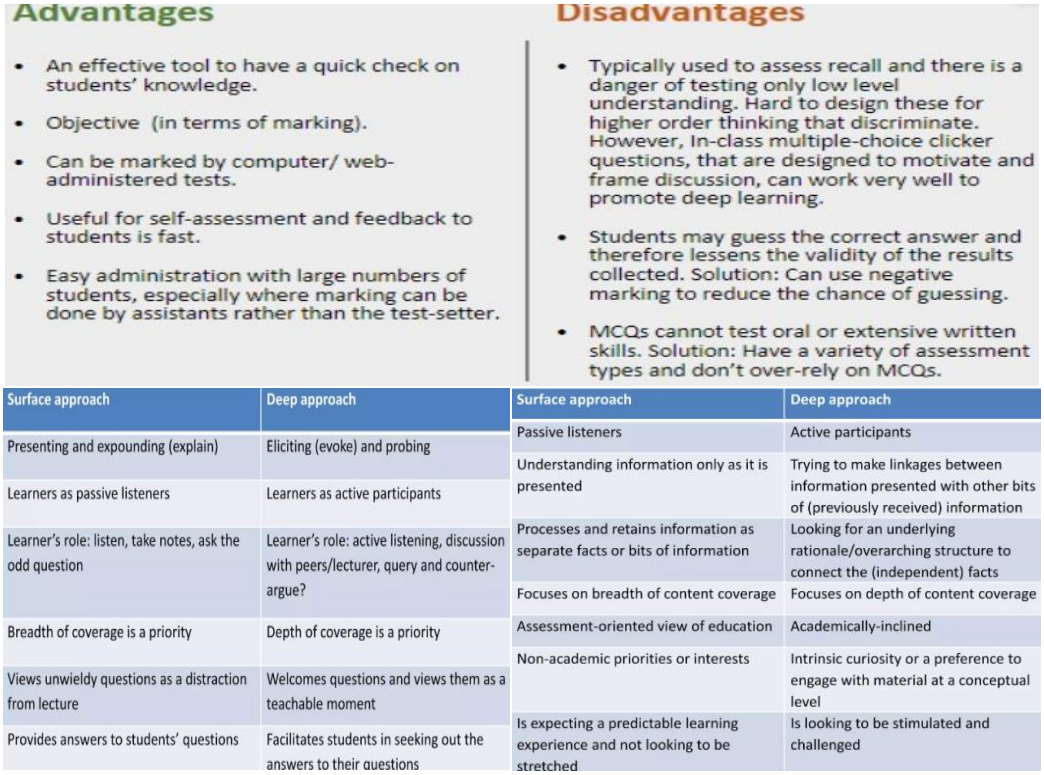


Fig. 6. Advantages&disadvantages for teachers&students faced with multiple choice exams, cf. Brame, 2013

Matching questions asks learners to pair clear items: definitions with specific concepts. For such items there are a variety of incorrect answers, difficult to predict. Hence it makes sense to

provide students with global feedback on the question and perhaps remediation if necessary. True-False questions require labelling statements, while multiple choice tasks want students to select the correct answer from a list of distractors, and matching questions ask students to pair the list of items provided. Such questions are quick &easy to create, and are marked automatically by the computer. Students need more time to answer them as compared to multiple choice or true/false items. Fill-in-the-blank questions are made of an incomplete statement for students to finish by typing one or two words that will make the statement correct.

Students have developed techniques for guessing correct answers. Here are some common guessing rules candidates have, and how to defeat them as test designer:

Guessing rule of thumb	**Test designers defeating the guessing rules**
Pick the longest answer	Make sure the longest answer is correct in 25% of the time if there are 4 alternatives or make all answers the same length
Pick the *b* option	Make each answer option the correct one an equal number of times
Never pick an answer with *always* or *never* in it	Make sure answers with *always* and *never* are correct as often as they are incorrect or avoid using these words
If 2 answers are express opposites, one is the correct answer	Offer opposites when neither is correct
If in doubt, guess	Increase the number of alternatives
Pick the scientific-sounding answer	Use scientific jargon in incorrect answers; make the simple answer correct
Pick a word related to the topic	Use terminology from the topic in distractors too.

Fig. 7. Students' guessing, combated by test design, cf. Nussbaum, 2008

Further tips for test designers:

- Avoid *all of the above/none of the above* as alternatives. Test-takers easily grasp that *all of the above* is always the correct answer, or that it is inserted due to lack of ideas for distractors.
- Give instructions. Some distractors may be correct under certain circumstances. Instruct students to select the best answer, rather than the right answer. It eliminates the debate about a distractor being correct under limited or isolated conditions.
- Be clear. Capitalize all limiting & directive words
- Grammar matters: check spelling and punctuation; avoid contractions; keep grammar consistency between the stem and alternatives.
- Reduce guessing. Avoid answering one question in the test by giving the answer elsewhere. Do not use extremes, complicated words, and unreasonable statements. Make options the same length and grammatically matched to the stem, mainly singular/plural. Avoid nonsensical distractors.
- Randomly distribute the correct answer among the positions.
- Avoid writing trick questions because they do not assess the student's knowledge.
- Letter the options to reduce confusion.

Advice for achieving objective tests:

- In true/false questions, check for qualifying words (all, most, some, none, always, usually, seldom, never, best, worst, highest, lowest, smallest, largest). Test for truth by substituting other members

of the series. If substitution makes a better statement, the statement is false; if substitution does not make a better statement, the statement is true.

- Watch for modifying/limiting phrases inserted into true-false questions. Test designers often add names, dates, places or other details to make a statement inaccurate.
- Be alert for multiple ideas/concepts in the same true-false statement. All parts of the statement must be true or the entire statement is false.
- Apply the similar tactics to answering true-false&multiple-choice items. The techniques function equally well for both, as multiple-choice items are just true-false questions arranged in groups.
- On matching exercises, work with only one column at a time. Match each item in that column against all items in the second column until getting a proper match.

Multiple-choice tests are a rapid and clear framework to assess students' grasp of a topic. It is a method of retrieval and practice, hence it assists in improving learners' retention and recall abilities. Subsequent to multiple-choice tests candidates may get false information.

Research has consistently shown that giving feed-back eliminates confusions and improves performance levels. An increasingly popular add-on to such exams is enforced self-confidence.

Research also shows that using such exam features makes test-takers enjoy assessment and develop effective study skills. They prove really useful in the acquisition of business engineering concepts in English, both in foreign language learning and in CLIL taught at academic levels.

Bibliography

1. Aronson, E. *Building empathy, compassion, and achievement in the jigsaw classroom,* in Improving Academic Achievement, Academic Press, 2002 pp. 209-225
2. Brame, CJ, *Writing Good Multiple Choice Test Questions,* Vanderbilt University Academic Press, 2013
3. Burton, J, Sudweeks, R, Merrill, F, Wood, B. *How to Prepare Better Multiple Choice Test Items: Guidelines for University Faculty*, OUP, 1991
4. Carlsmith, K, P*ersuasive example of collaborative learning* Teaching of Psychology, 2002, 29, 132-135
5. Cismaş SC, *Culture-Fair Assessment of Foreign Language Skills in Engineering Education* Recent Advances in e-Activities, Information Security&Privacy, Computer Engineering Series 2009, pp. 263-267
6. Cismaş, SC, *Abordarea unitară a celor 4 deprinderi de limbă străină si standarde specifice* DPPD Univ Bacău, Revista Studii şi Cercetări Ştiinţifice Seria Stiinţe Socio-Umane, nr 3/2000, pp. 114 – 118
7. Cismaş, S.C, *Being an ESP Teacher* Prosper Newsletter (Project Special Purposes English in Romania), issue 11, October 1999, The British Council, Bucharest, pp. 60-61
8. Cismaş, SC, *Brief History of Educational Objectives&Specific Impact on Foreign Language Acquisition in Engineering* Recent Advances in e-Activities, Computer Engineering Series pp. 257-262
9. Cismaş, SC *Developing Creativity&Problem Solving Skills by English Grammar&Vocabulary Activities in Engineering* Proceedings 8th WSEAS Int'l Conference on Education, Recent Advances in Computer Engineering, Genoa University 2009, pp. 161-166
10. Cismaş, S.C, *Standarde de competenţă pentru profesorii de limbi străine,* Universitatea Politehnica din Timişoara, Analele Facultăţii de inginerie din Hunedoara, Anul 3 Fascicula 7, 2001 pp. 224-227
11. Cismaş, S.C., *Authentic Assessments,* Symposium on Advanced Topics in Electrical Engineering, ATEE 2000, Bucharest Polytechnic University http://iem.pub.ro/ro/evenimente/ATEE-2000/program_ATEE.htm
12. Cismaş, SC, *Corelarea evaluarii cu procesul de predare-invatare in engleza tehnica*, vol Limba, cultura, civilizatie, Directii de optimizare, a VI-a conferinta internationala, POLITEHNICA Press 2012 pp. 141-149
13. Cismaş, S.C., *Performance Evaluation in Foreign Language e-Learning* 9th WSEAS International Conference on Distance Learning&Web Engineering in Science, 2009 pp. 110-115
14. Cismaş, SC, *Test Anxiety and Motivational Incentives in Web-Based Learning* Proceedings 9th WSEAS International Conference on Distance Learning&Web Engineering in Science, 2009 pp. 77-82
15. Cismaş, S.C., *Templates for Scientific Writing in Homework &Exams for Engineering Students,* Latest Trends on Engineering Education, 2010, 7th WSEAS International Conference, pp. 193-198

16. Cismaş, SC, *Limbile străine in invăţământul superior tehnic. Standarde curriculare pentru studenţi*, Universitatea Politehnica Timişoara, Analele Facultăţii de Inginerie, An 3 Fascicula 7 2001 pp. 228-233
17. Cortright, R, Collins, H, Rodenbaugh, D, *Student retention of course content improved by collaborative group testing.* Advanced Education, 2003, 27, 102-108.
18. Dihoff, R, Brosvic, G, Epstein, M, *Provision of feedback during preparation for academic testing: Learning enhanced by immediate not delayed feedback.* 2004, The Psychological Record, 54, 207-231.
19. Dunn, I, Morgan C, *New Directions in Traditional and Online Assessment,* London, Routledge, 2004
20. Ebel, R, Frisbie, D. *Essentials of educational measurement* Englewood Cliffs, NJ: Prentice-Hall, 1986
21. Haladyna T, Downing S, *Taxonomy of multiple-choice* Applied Measurement in Education 1989, 2(1), 37-50.
22. Haladyna, T, Downing, S, *Developing and validating multiple-choice test items,* Erlbaum Assoc, 1999
23. Haladyna, T, Downing, S, *Validity of multiple-choice* Applied Measurement in Education, 1989, 2(1), 51-78.
24. Kennette, L, McGuckin, *D, Using the Immediate Feedback Assessment Technique for non-assessments: Student perceptions and performance,* 2018, Psychology Teaching Review 24, 66-71
25. Koh, A, Lee, C, *The learning benefits of teaching,* Applied Cognitive Psychology, 2018, 32, 401-410.
26. Laal, M, *Benefits of collaborative learning,* Procedia Social&Behavioral Sciences, 31, 486-490, 2012
27. Morrison, S, Free, K, *Writing multiple-choice test items to promote & measure critical thinking,* Journal of Nursing Education 40: 17-24, 2001
28. Nussbaum, E, *Collaborative discourse, argumentation, and learning: Preface and literature review,* Contemporary Educational Psychology, 33, 345-359, 2008
29. Raadt, M. *Tips on Taking Multiple Choice Tests*. Social Psychology Network, 2018