DESIGNING GREAT ASSESSMENT

The case for using multiple choice questions for accurate assessment

Evidence Based Education

WHY IS IT IMPORTANT TO HAVE GREAT ASSESSMENT?

Effective assessment is a process designed to generate inferences, or actionable meanings. It aims to help teachers achieve clearly-defined goals that they couldn't reach without it.

Taking a test often does more than assess knowledge; tests can also provide opportunities for learning (Roediger & Karpicke, 2006). For example, a well-planned assessment could give you the opportunity to provide high-quality feedback without adding to your own workload.

One way to ensure your assessment is effective is to plan for it to be a strong bridge between teaching and learning (Wiliam, 2013). When you have great assessment, the process of ascertaining where the learner is, where they need to be, and how to get them there becomes much easier!

WHY ARE MULTIPLE CHOICE QUESTIONS A GREAT ASSESSMENT TOOL?

Of course, there are many ways to assess. In this guide, we focus specifically on the use of multiple choice questions (MCQs).

Typically, teachers use MCQs for testing facts and so-called lower-order thinking skills to test if children remember and understand facts (Tarrant et al., 2006; Zimmaro, 2010). However, when MCQs are purposeful and well-planned, they can spark higher- level learning, and create a desirable difficulty, positively increasing the value of the assessment (Coe, 2013).

As with any educational tool, you should make a judgement about whether MCQs suit your particular purpose before deciding whether to use them.

THE BENEFITS AND LIMITATIONS

BENEFITS

- They produce more reliable information than other types of test such as open- ended questions, as they're objectively scored (Burton et al., 1991).
- The student can answer more multiple choice questions than open-ended questions, covering a broader, more representative subject base[JD1] (Zimmaro, 2010; Burton et al., 1991)..
- They're quick to mark and quick to complete! (Christoadoulou, 2016)!.
- A well- crafted MCQ will allow the teacher to identify misconceptions easily and in the same amount of time (Christodoulou, 2016)

LIMITATIONS

- They can be susceptible to guessing, (Burton et al. 1991), although it's unlikely that a student will ace a test adopting this technique (Christadoulou, 2016).
- They cannot measure production skills like writing ability, creativity, or articulation of thought (Zimmaro, 2010).
- They are not quick to construct, and good MCQ construction takes time (Zimmaro, 2010).
- MCQ scores are influenced by risk-taking traits. Boys are more likely to guess than girls, and risk -taking is also associated with certain cultures and socio-economic background (Ben-Shakhar and Sinai, 1991).

DESIGNING A MULTIPLE-CHOICE TEST

Once you've considered the evidence, you need to think about how you design a multiple –choice test.

Before writing the questions in such a test, you need to be confident that you're not just testing a small part of the subject area based on the key learning objectives in your curriculum.

One way to do this is to create a table based on constructs, or learning objectives, in your schemes of work. This is a systematic approach to help identify both the learning objectives you need to cover and the complexity of understanding that you want to assess.

In the table below, the teacher has used the blueprint to design a multiple– choice test that assesses if learners can identify and interpret the use of present tense verbs. It might look complex at first but follow the step-by-step guide to demystify the process.

A BLUEPRINT FOR A FRENCH KEY STAGE 3 ASSESSMENT

Learning Objective	Identify/ Name	Interpret/ Describe/ List	Weight	Target no. questions (per learning objective)
Regular present tense er verbs	2	2	40%	4
Irregular present tense er verbs	1	1	20%	2
Present tense ir verbs	1	1	20%	2
Present tense re verbs	1	1	20%	2
Weight	50%	50%	100%	
Target no. questions (per skill)	5	5		

The weight and target number of questions are entirely up to you and should be based on your learners' needs. This blueprint reflects an assessment with a purpose to assess a learner's base understanding of key grammatical concepts, rather than their production skills.

Remember, this blueprint can be used for planning and weighting any sort of assessment – from the informal, in-class hinge questions you might ask verbally, right through to the written end-of-unit test. So, here's how to make your own.

STEPS TO CREATING A BLUEPRINT

Step 1: Decide how many questions are needed to achieve the test's purpose. The number of questions should be created to fit into the optimum amount of time you are willing to spend on the assessment, whether that's seconds or hours. A test of ten questions gives a fair representation of the constructs to be assessed, with more questions weighted towards a construct which is more fundamental than others.

Step 2: Decide what type of skills are useful for accessing the content of the assessment, then write them across the x axis. These skills should reflect the depth of understanding based on a taxonomy (e.g. Bloom's, or SOLO). Here, the purpose is to assess the learner's base understanding of key present tense grammatical concepts rather than their production skills. Therefore, the constructs chosen are identify, interpret, and analyse.

Step 3: Add what you want to assess. The items along the y axis are your learning objectives, or constructs. These should be taken from the scheme of work derived from the curriculum. The list should only include elements you want to test because you can do something with the information generated; anything further is irrelevant for the purpose of this test. This teacher wanted focus on the understanding of 'er' verbs, as that had been a focus in previous lessons, while also checking other regular verbs.

Step 4: Enter a percentage indicating how much weight you want each learning objective and skill descriptor to have. This depends on the focus that you need for your work. The main thing here is to make sure that the weights total 100%!

[adapted from Rodriguez and Albano, 2017]

DESIGNING GREAT MULTIPLE –CHOICE QUESTIONS
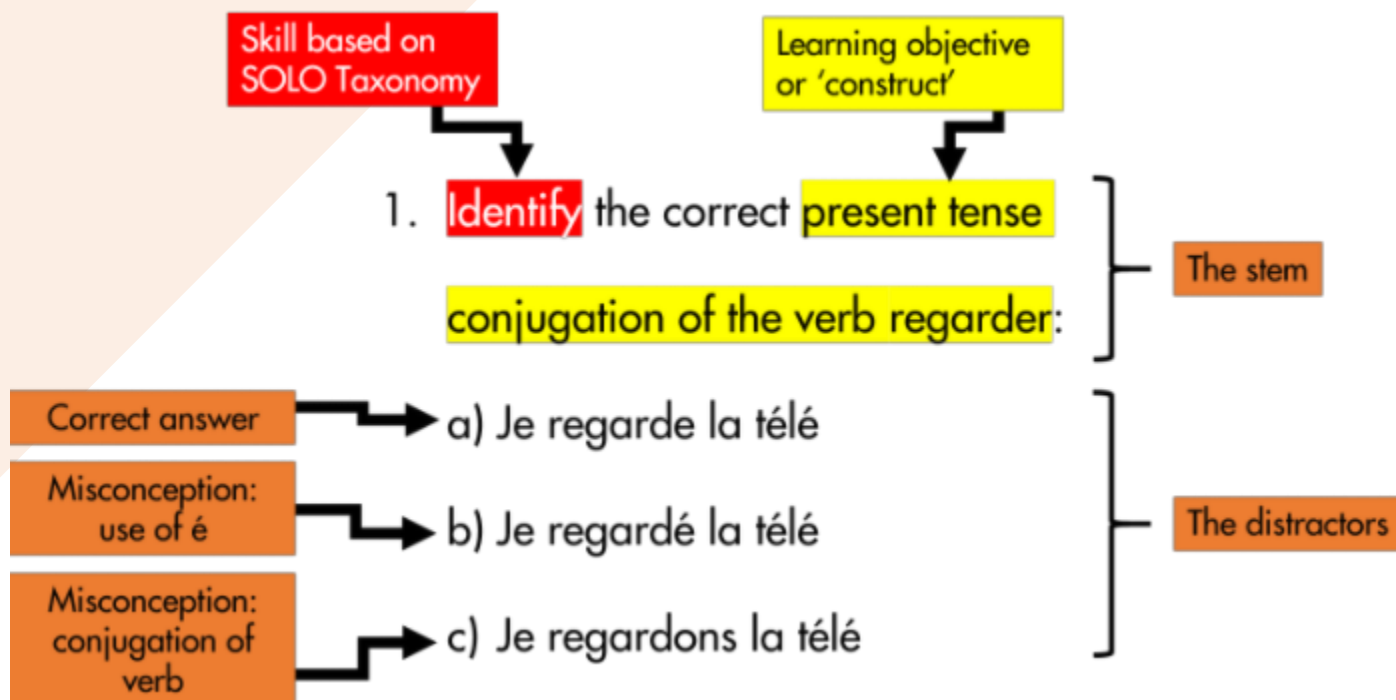
- MCQs consist of two parts: the stem and the alternatives.
- The stem should be meaningful (Brame, 2013) and should include a question which stimulates answers and focuses on specific learning outcomes.
- The alternatives include the correct answer, or key, and alternatives.
- The alternatives should be similar in content (Boland et al., 2010; Burton et al., 1991).
- The correct answer should not be longer or more detailed, as this can sometimes be interpreted as a clue (Tarrant & Ware, 2008).
- The vocabulary should be accessible (Burton et al., 1991; Moreno et al., 2006).
- The questions should be independent of one another: the next question should not require answering the previous one correctly (Brame, 2013).
- Questions should avoid negative phrasing, e.g. "which of these is not a…" (Burton et al., 1991).
- Avoid using words like 'never', 'always', 'usually' (Boland et al., 2010).

MULTIPLE-CHOICE CHECKLIST

Checklist	Yes	No
Before writing the items, the learning goals were weighted based on a test blueprint		
Each question assesses only one clear learning objective		
Each question is independent from one another		
The assessment does not use unnecessarily hard vocabulary		
There are no opinion-based questions		
The easy questions are placed at the beginning of the test		
Use of letters in front of the alternatives, not numbers		
All the stems are phrased in the shortest possible way		
All the stems do not include irrelevant content		
The stems are not negatively phrased		
The stems do not give clues and exclude some of the alternatives		
The alternatives are mutually exclusive and do not overlap		
All the alternatives are presented in a logical order		
All the alternatives are phrased in a similar way		
There is only one correct answer in each question		
The right option does not offer more details, nor is it longer		
All the distractors are plausible		
The distractors are based on students' misconceptions		

NEXT STEPS

When analysing student response data, the effective use of alternatives will help you identify misconceptions in your pupils' learning, which is as useful as knowing what they answered correctly. If a distractor is chosen by less than 5% of the students, it might be advisable to reconsider whether this distractor should be removed or amended (Tarrant et al., 2009; Towns, 2014).

Once you have identified specific misconceptions, you can direct your pupils to the relevant supplementary exercise, reviewing previous learning or moving on to a 'stretch and challenge' task for those who excelled. Using assessment for formative assessment in this way makes it meaningful for the learner and a much quicker and more reliable data collection exercise for you.

The use of MCQs as retrieval questions, hinge questions, tests, and in assessment for learning can be explored further through engaging with Evidence Based Education's Assessment Academy. Assessment Academy bridges the gap between academia and the classroom to help teachers assess better. Find out more at evidencebased.education

GLOSSARY OF TERMS

Alternatives: the options provided by multiple-choice questions, including both the right and the wrong answers.

Blueprint (assessment blueprint): a written plan/map of the assessment specification which matches the main topics to cognitive skills and number of questions assessed.

Domain: area

Distractors: the wrong alternatives provided as options in a multiple-choice question. Item: question in a test

Key: the correct option in a multiple-choice question

Key balancing: Systematic techniques used to improve the patterns that a correct answer appears in a test.

Misconception: a wrong view on a topic because there is flawed understanding.

Stem: It is the part of the multiple-choice question which sets the question and demands a response.

Testwiseness: the behaviour that allows test takers to maximize their test score even though they do not know the answer to a multiple-choice item. This might happen due to guessing or deducing a correct answer based on cues in the question (Downing, 2002).

REFERENCES

Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: the role of differential guessing tendencies. *Journal of Educational Measurement*, 28(1), 23-35.

Boland, R., Lester, N., & Williams, E. (2010). Writing multiple-choice questions. *Academic Psychiatry*, 34(4), 310-316.

Brame, C. (2013). Writing good multiple-choice test questions. Retrieved from https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions
Burton, S. J., Sudweeks, R. R., Merrill, P. G., & Wood, B. (1991). *How to Prepare Better MultipleChoice Test Items: Guidelines for University Faculty*. Provo, UT: Brigham Young University Testing Services and The Department of Instructional Science.

Christodoulou, D. (2016). *Making Good Progress? The future of Assessment for Learning*. Oxford, England: Oxford University Press.

Coe, R. (2013). [online] Cem.org. Available at: http://www.cem.org/attachments/publications/ImprovingEducation2013.pdf [Accessed 13 Mar. 2019]. Moreno, R., Martínez, R. J., & Muñiz, J. (2006). New guidelines for developing multiple-choice items.

Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2(2), 65-72. Rodriguez, M. C. (2005). *Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research*. Educational Measurement: Issues and Practice, 24(2), 3-13.

Roediger, H. and Karpicke, J. (2006). *Test-Enhanced Learning*. [online] St Louis: Washington University, pp.249-255. Available at: http://learninglab.psych.purdue.edu/downloads/2006_Roediger_Karpicke_PsychSci.pdf [Accessed 13 Mar. 2019].

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments. *Nurse Education in Practice*, 6(6), 354-363.

Towns, M. H. (2014). Guide to Developing High-Quality, Reliable, and Valid Multiple-Choice Assessments. *Journal of Chemical Education*, 91(9), 1426-1431.

Wiliam, D. (2006). *Assessment: The Bridge between Teaching and Learning*. Voices from the Middle, 21(2), pp.15-20.

Zimmaro, D. M. (2010). Writing Good Multiple-choice Exams. Retrieved from: http://ctl.utexas.edu/assets/Evaluation--Assessment/Writing-Good-Multiple-Choice-Exams-04-28-10.pdf