p-ISSN: 2184-044X e-ISSN: 2184-1489 ISBN: 978-989-35106-9-8 © 2024
https://doi.org/10.36315/2024v1end024

ON SCORING COMPETENCE

Magne V. Aarset[1], & Stig Ytterstad[2]
[1]Department of Ocean Operations and Civil Engineering,
Norwegian University of Science and Technology (Norway)
[2]Department of Leadership and Organizational Behaviour,
Norwegian School of Management BI (Norway)

Abstract

The overall objective of any learning process is (at least) threefold; to increase a learner's competence, confidence, and learning ability. The totality of the learning process should improve the learner's social ability to learn from the whole process on how to be able to have stylistic flexibility and, by that, use her learning preferences to handle a test optimally, so her learning preferences make the best fit to the test. Here, we are focusing on introducing scoring techniques to assess a learner's competence concerning a specific subject from responses to multiple-choice tests, thereby being able to provide autonomous adapted feedback within the constraints of increasing learner awareness of own learning style and thereby strengthen confidence and learning ability. How a learner decides to act is essential for acquiring and maintaining satisfactory situational and learning style awareness. Therefore, information of the test and evaluation design is mandatory during the learning process. Feedback to the learner should ideally include the history characterizing the learning process so far, an estimate of the current state of the learning process, and a forecast of both the expected, an optimistic and for some learners, maybe even a worst-case scenario of the learning outcome. Feedback should be formed to strengthen the learner's situational awareness regarding her learning style, thereby reducing the probability of human failure, motivating the learner (e.g., by experiencing mastery), and suggesting beneficial learning strategies.

Keywords: *Assessments in education, scoring competence, multiple-choice tests.*

1. Introduction

During the last decades learning has again become a key topic. This time not only at learning institutions, but also in political and economic contexts. One reason for this is that a high level of competence (i.e., knowledge, understanding, and/or skills) in nations, organizations and individuals are considered both a necessity and a crucial competitive advantage in the present knowledge society and the globalized market (Aarset & Johannessen, 2022).

The overall objective of any learning process is (at least) threefold. It is to increase a learner's competence, confidence, and learning ability to improve the learner's ability for analytical and critical thinking and social skills, and thereby preparing them for society (Aarset & Johannessen, 2022).

Here, we are focusing on introducing scoring techniques to assess a learner's competence with respect to a specific subject from responses to assessments, for thereby being able to provide adapted feedback. Such an assessment is thus a quantitative or qualitative estimate to indicate the competence level of the learner. For this being successful, the assessment score should, as close as possible, correspond to the "true level" of competence. How to measure this competence has for ages been discussed and debated without any clear conclusion being accepted in the literature, though.

A learner's understanding of her own learning process and her own competence (i.e., her situational and learning style awareness), will control her attention and influence how she decides to act. By learning style awareness, we mean the conscious dynamic reflection of the learning process by an individual learner. Situational awareness is often understood as a summary, a mental "picture", of what's going on in our world, and by some, also including the process of acquiring this summary. People make bad judgments when their overall understanding of what's going on is insufficient. This goes also for the learning process.

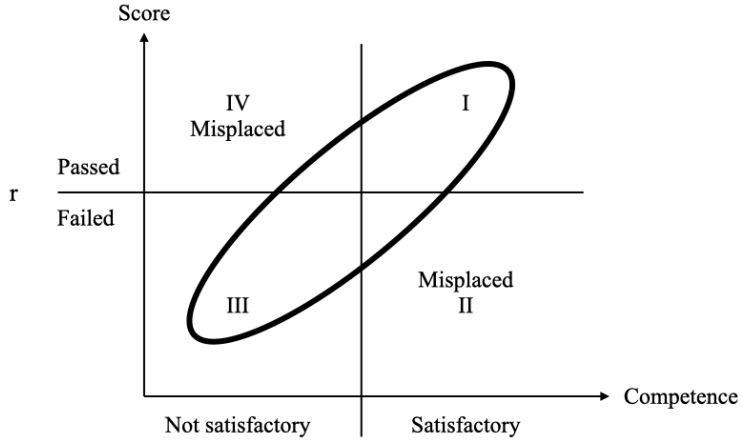
To acquire and maintain satisfactory situational and learning style awareness during a learning process, feedback to the learner needs to be provided. For this feedback being beneficial, learners must know the premises of the test clearly in advance so that the test can become part of the learning process.

Knowledge of the framework of the test will strengthen the learning process associated with it. Feedback should be formed to strengthen the learner's situational and learnings style awareness, and thereby reduce the probability of human failure in addition to motivate the learner (e.g., by experiencing mastery) and suggest beneficial learning strategies.

The objective of this manuscript is to suggest a way of scoring competence and provide this as feedback to an individual learner during a learning process. Even though competence clearly is a multidimensional concept, competence is commonly measured on a scale from 0 – 100 (%), where 0 means "*Non-competent*" (or *Incompetent*) and 100 "*Fully competent*". This attribute is typically measured by the metric "Percentage of correct answers" on a multiple-choice test. Here, we will use the scale [-1, 1], or a sub-set of this interval, when scoring competence.

Basically, such assessments seek to discriminate between learners who have a satisfactory competence within a specific field and those who hasn't. Thus, the goal is to discriminate between learners in what we have called group I and III in figure 1 below, and simultaneously (mis-)placing as few as possible learners into group II or IV.

Figure 1. Classification of learners.



All teachers (facilitators) hope learners end up in group I, having both a thorough competence of the field they are supposed to learn and being able to communicate this in an assessment by for example passing a test. Simultaneously, it's important for a teacher (and e.g., a learning institution) to identify the learners without satisfactory competence. In a good assessment, the vast majority of learners will either belong to group I or III.

In group IV learners have managed to discover the correct answers to enough assessment items without satisfactory competence. They typically accomplish this by guessing, or by cramming and memorizing "type problems" and procedures. To (partly) rectify this we might require a high total score for passing the assessment (indicated with a large *r* on figure 1).

Requiring a high total score for passing an assessment will on the other hand increase the problem regarding learners in group II that have managed to acquire a satisfactory level of competence in the specified field, but still fail the test. Thus, the assessment of their accomplishment is not measured correctly and the validity with respect to measuring competence is not satisfactory. To (partly) rectify for this we might require a low total score for passing the assessment (indicated with a small *r* on figure 1), contrary to what was just suggested to minimize the probability of misplacing a learner into group IV.

Measuring knowledge, understanding and skills are difficult, if not impossible, in both a classroom setting and in a "real life" situation. As testing competence by automatic test procedures such as multiple-choice tests, sometimes even presented by an AI system, has become popular it is of fundamental importance to establish good testing procedures.

Another concern is that learners can answer correctly through guessing, making it impossible to distinguish between correct answers based on competence versus pure luck. Without the ability to solve a particular item, learners may gain marks by guessing and thereby introduce a random factor into test scores that lowers reliability and validity.

2. Scoring competence

A *Multiple-choice* test where the learner is instructed to identify the one correct statement among k_j statements in J items (i.e., one correct statement and k_j -1 distractors for each item, j=1, …, J) is a common type of scoring competence in assessments. Traditionally, multiple choice tests have been scored

using a conventional *Number Right* scoring method. Correct answers are scored with a value of 1, incorrect answers and omitted answers with a value of zero, and the sum of the scores for correct responses is the score of the assessment.

Now, let X_j (j = 1, … , J) be a stochastic variable characterizing the score for answering item (question) number *j*, and $X = \sum_{j=1}^{J} X_j$ the sum of scores, where

$$X_j = \begin{cases} 1 & if\ correct \\ 0 & if\ incorrect. \\ 0 & if\ omitted \end{cases}$$

In a multiple-choice test with one correct answer out of k_j alternatives on item *j* (*j* = 1, … , J), the expected value of each X_j will be positive if a learner doesn't know the correct answer and is just guessing.

$$EX_j = 1 \cdot \frac{1}{k_j} + 0 \cdot \frac{k_j - 1}{k_j} = \frac{1}{k_j}.$$

So basically, it is expected to be beneficial for a learner to guess if she doesn't know the correct answer. And if she does, it is impossible for a teacher (facilitator) to distinguish between competence and pure luck.

Therefore, various scoring formulas have been presented to correct for guessing. The *Rights Minus Wrongs* (or *Negative Marking*) scoring method penalizes the learner for incorrect responses. The fundamental idea behind this scoring method is that learners acknowledge they will lose marks for incorrect answers and become discouraged to guess. This is expected to increase test reliability and validity because the test score will more truly reflect the learner's competence.

If the learner decides to guess with such a scoring technique, the expected score is

$$EX_j = 1 \cdot \frac{1}{k_j} + (-1) \cdot \frac{k_j - 1}{k_j} = -\frac{k_j - 2}{k_j}.$$

With more than two alternatives per item ($k_j > 2$), a learner with no knowledge is discouraged from guessing because the expected score will be negative. And the more distractors, the less will the expected score become and the more the learner is discouraged from guessing. But this may be too much discouraging, introducing a level of risk acceptance among the learners as a bias when assessing competence.

A favourable characteristic with a scoring method may be that the expected score should be zero if a learner is guessing an answer at random. For this to happen when the reward for getting the question correct is 1, the penalty for an incorrect answer should be $-1/(k_j-1)$, where k_j is the number of alternatives. Omitted items is given a score 0.

$$X_j = \begin{cases} 1 & if\ correct \\ \frac{-1}{k_j - 1} & if\ incorrect. \\ 0 & if\ omitted \end{cases}$$

Now, in a multiple-choice test with k_j alternatives where only one alternative is correct, the expected score on such an item is

$$EX_j = 1 \cdot \frac{1}{k_j} + \left(-\frac{1}{k_j - 1}\right) \cdot \frac{k_j - 1}{k_j} = 0.$$

This kind of neutrality with respect to guessing may be beneficial. If the learner is able to eliminate one or several alternatives, though, it will be beneficial to guess.

3. Partial competence and partial knowledge

Learners may be able to determine that some of the choice options are (clearly) incorrect even though they cannot identify the correct answer. Being able to eliminate some incorrect alternatives (distractors) is reflecting what is called "*partial competence*" and "*partial knowledge*".

If a learner eliminates *one* distractor correctly in a multiple-choice test as mentioned above, the expected score is

$$EX_j = 1 \cdot \frac{1}{k_j - 1} + \left(-\frac{1}{k_j - 1}\right) \cdot \frac{k_j - 2}{k_j - 1} = \frac{1}{(k_j - 1)(k_j - 1)}.$$

If the learner eliminates *two* distractors correctly, the expected score is

$$EX_j = 1 \cdot \frac{1}{k_j - 2} + \left(-\frac{1}{k_j - 1}\right) \cdot \frac{k_j - 3}{k_j - 2} = \frac{2}{(k_j - 1)(k_j - 2)}.$$

And so on.

Table 1. Expected score after correctly elimination of distractors.

EX_j		Correctly eliminated distractors				
		0	1	2	3	4
Number of alternatives, k_j	2	0	1			
	3	0	**.25**	1		
	4	0	**.11**	**.33**	1	
	5	0	**.06**	**.17**	**.38**	1
	6	0	**.04**	**.10**	**.20**	**.40**

Table 1 illustrates that with the scoring method mentioned above, the effect of guessing is neutral, and a learner who is not able to identify any distractors (i.e., a non-competent learner) will have an expected score equal to 0. If a learner manages to correctly eliminate some distractors, though, guessing will have a positive expected effect. To avoid this positive effect by guessing, it is possible that an effective penalty that discourages guessing should exceed the standard penalty of $-1/(k_j - 1)$, but by implementing such negative marking, this will reflect the learners' answering strategies and risk-taking behaviour instead of actual competence. In fact, we see it as beneficial that for a learner able to eliminate some distractors (i.e., a learner with partial knowledge) the expected score is positive (> 0).

Therefore, we define a learner

- to have *full knowledge* if $E(X_j) = 1$
- to have *partial knowledge* if $0 < E(X_j) < 1$, because of incomplete information, lack of confidence, or both
- to be *non-competent / incompetent* if $E(X_j) = 0$, i.e., the learner is guessing or omitting to answer.

Frary (1989) introduced *partial-credit scoring methods*, and a variety of different partial-credit scoring methods are presented in the literature. We distinguish between three main formats:

- The *liberal multiple-choice test* allows learners to select more than one answer to a question if they feel uncertain which alternative is correct (Bush, 2001).
- In *elimination testing* learners are instructed to cross out all alternatives they consider to be incorrect (Kurz, 1999).
- In *confidence weighting* learners are instructed to indicate what they believe is the correct answer and how confident they are about their choice (Kurz, 1999).

Still an alternative model, proposed by Traub et al. (1969), rewards a learner for not guessing by awarding points for omitting items rather than penalizing for incorrect responses. This presents a psychological advantage since it rewards the desired behaviour rather than penalizing undesirable behaviour. Learners do not feel threatened by receiving a reward for skipping items, as compared to receiving a penalty for incorrect responses.

Hattie (2012) writes that to make learning and teaching visible, it is essential that the teacher also acts as an evaluator who activates and manages to embed good learning strategies in the students. It makes one distinguish between surface and in-depth knowledge and understand the conceptual difference between them. The teacher must have insight into learning styles so that she can communicate this to the learners and help them understand the connection between what they are learning and the feedback they will receive through multiple-choice tests. One of the points about excellence in education is that both the teacher and the learner need to know the intentions of what is going to be learned and what criteria there are for evaluation. This means that they must know how to meet these criteria and thus take the steps to fill the gap between the learner's current knowledge and the level needed to pass the test. Good teachers, therefore, introduce to the learners how to learn and arrange for them to be prepared for the type of test they receive. It is essential that the learning process has a clear intention and that the success criteria are explicitly known among the learners during the learning process. Therefore, the learners must understand the necessity of being able to answer the test questions precisely.

Furthermore, the learners must have some relationship to whether their answers meet the success criteria of the multiple-choice tests and that they can distinguish between correct and incorrect answers. The learners must know why the answer is correct and, if necessary, also know what other information it will be necessary to have to meet the requirements of the test. Feedforward is a relevant term in this context. It is the process by which one imagines that something is going to happen, or could happen, sometime after the event and investigates whether the performance held up. It is thus an understanding or an assumption that something particular will happen with a subsequent investigation of whether what was done to realize the performance also happened. This will make students aware of what kind of calibration they should have in relation to the test they are going to conduct. This will help to strengthen the test's success rate (Hermansen, 2003).

Zhang et al. (2012) writes that a thinking style or learning style is the preferred way of thinking. It is not a trait but instead how we use our properties. We do not have a style, but a profile of styles. Zhang et al. (2012) discuss thinking styles and instruction methods and writes that thought-based questions are closest to the two learning styles, the Judicial and Legislative ways of thinking. They write that memory-based analysis and time allocation are the skills used on short answer assignments or multiple-choice tests. One of the ways one learns is to work for oneself, and that is central. This affects the Executive, Local, Judicial, and Hierarchical Learning styles. It emphasizes the importance of learners being introduced to the main features of their own way of thinking and how it can be utilized in such a test situation. In the "Handbook of Intellectual Styles" (Zhang et al., 2012), they write, referencing Furnham, Swami, Arteche, and Chamorro-Premuzic (2008), that what is called "surface learners" prefer multiple-choice and group work rather than being assessed based on essays or dissertations. Those who are "deep learners" need to practice mastering the multiple-choice form well to get usable results. For this to happen, the teacher must include a review of different learning styles in their course and evaluation design so that the students receive an optimal and fair outcome (Biggs & Tang, 2011).

4. Conclusion

Depending on how strong a learner's preferred learning style is on the different styles, it will be beneficial for the student to have a conscious relationship with her learning styles. Practicing and adapting relevant learning styles before such a test situation will allow the student to get the most out of both the learning process and its evaluation (Zhang et al., 2012).

We suggest introducing a scoring of competence where the effect of guessing is neutral, but if a learner manages to correctly eliminate some alternatives before guessing, guessing should have a positive expected effect reflecting "*partial knowledge*".

It seems too spendable to reward a learner for managing to identify less than the expected correctly identified alternatives if she is just guessing. We therefore suggest introducing a penalty when a learner is scoring less than the expected score when guessing. Introducing and explaining this evaluation method will strengthen her awareness of the necessity of having precise and thorough knowledge before the test.

References

Aarset, M. V., & Johannessen, L. K. (2022). On Distributed Cognition While Designing an AI System for Adapted Learning. *Frontiers in Artificial Intelligence, 5*, 910630. Doi: 10.3389/frai.2022.910630

Biggs, J., Tang, C., & Kennedy, G. (2022). *Teaching for quality learning at university* (5 ed.). London: McGraw-Hill Education.

Bush, M. (2001). A multiple choice test that rewards partial knowledge. *Journal of Further and Higher education*, *25*(2), 157-163.

Frary, R. B. (1989). Partial-Credit Scoring Methods for Multiple-Choice Tests. *Applied Measurement in Education, 2*(1), 79-96. doi:10.1207/s15324818ame0201_5

Furnham, A., Swami, V., Arteche, A., & Chamorro-Premuzic, T. (2008). Cognitive ability, learning approaches and personality correlates of general knowledge. *Educational Psychology*, *28*(4), 427-437.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hermansen, M. (2003). *Omlæring*. Klim.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. New Jersey: FT Press.

Kurz, T. B. (1999). *A Review of Scoring Algorithms for Multiple-Choice Tests*. Retrieved from https://files.eric.ed.gov/fulltext/ED428076.pdf

Traub, R. E., Hambleton, R. K., & Singh, B. (1969). Effects of Promised Reward and Threatened Penalty on Performance of a Multiple-Choice Vocabulary Test. *Educational and Psychological Measurement, 29*(4). Doi: 10.1177/001316446902900410

Zhang, L. F, Sternberg, R. J., & Rayner, S. (Eds.). (2011). *Handbook of intellectual styles: Preferences in cognition, learning, and thinking*. Springer Publishing Company.