Chapter # 1

ON SCORING COMPETENCE

Magne V. Aarset[1] & Stig Ytterstad[2]
[1]*Norwegian University of Science and Technology, Norway*
[2]*BI Norwegian Business School, 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 different kinds of multiple-choice tests, thereby being able to provide autonomous adapted feedback within the constraints of increasing learner awareness of own learning style and 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 should be provided to the learner during the learning process.

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 reglobalized 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 of 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 (Salmon, Stanton, Walker, & Jenkins, 2009). 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 learning 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. 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.

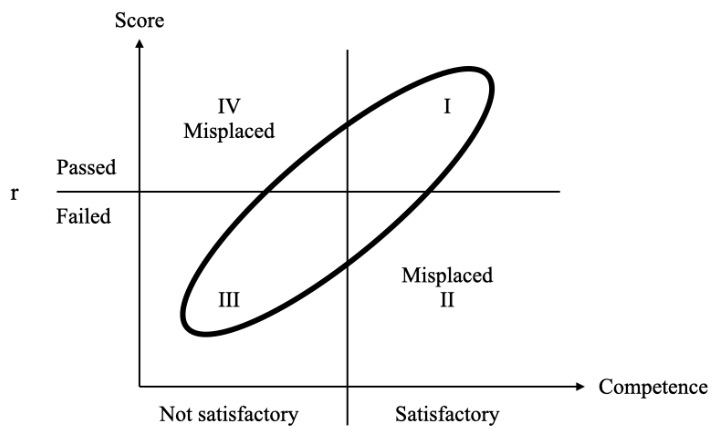
2. BACKGROUND

2.1. Introducing a Utility Function for Scoring Competence

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 $[-\infty, 1]$, 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 groups 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.

Learners placed in groups II and IV are misplaced. 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).

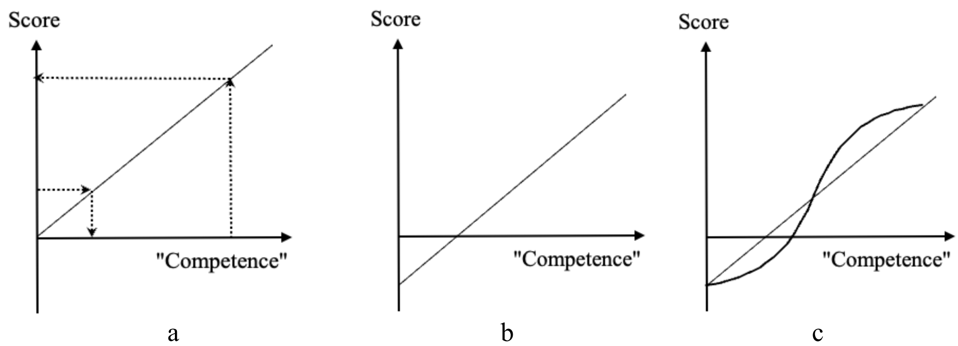
Requiring a high total score for passing an assessment will on the other hand increase the problem regarding learners in group II that have managed to acquire a satisfactory level of competence in the specified field, but still fail the test. Thus, the assessment of their accomplishment is not measured 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.

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.

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.

As there are several issues with respect to scoring competence, we will address some characteristics such a utility (scoring) function from a set of possible outcomes of a test should possess. As mentioned above, it's common to introduce a linear utility function from the interval [0, 100] to for example the interval [0, 1] as illustrated on figure 2a. We will in the following argue that introducing some kind of penalty, i.e. negative scoring, if the quality of the learner's answering is "poor" as illustrated on figure 2b might be beneficial. Furthermore, we suggest that it might be beneficial to let the utility function be S-shaped as illustrated on figure 2c. This is in line with scoring as performed in Item Response theory (IRT) (Hambleton, Swaminathan, & Rogers, 1991).

Figure 2.
a, b and c
Examples of utility functions for scoring competence.



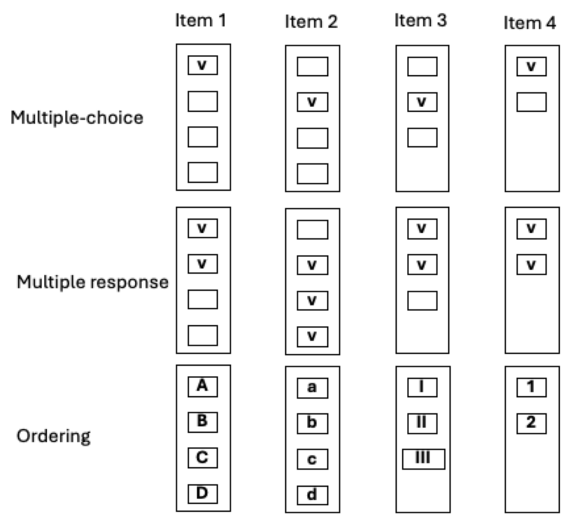
3. SCORING COMPETENCE

3.1. Common Techniques of Scoring Competence in Assessments

Some common techniques of scoring competence in assessments are (see figure 3);

- *Multiple-choice,* where the learner is instructed to identify the one correct statement out of k_j statements in each of J items (i.e., one correct statement ($c_j = 1$) and k_j - c_j distractors for each item, j = 1, … , J). $k_j > c_j = 1$.
- *True/False,* where the learner is instructed to identify the one correct statement out of two in each of J items. (This can be analyzed as *Multiple-choice* with $k_j = 2$ and $c_j = 1$.)
- *Fill in,* where the learner is instructed to select one word out of k_j words in each of J items. (This can be analyzed as *Multiple-choice*.)
- *Multiple response,* where the learner is instructed to identify the c_j (>1) correct statements out of k_j statements in each of J items. When the learner is informed about the value of c_j, we assume $k_j > c_j > 1$ and when the learner is not informed, we assume $k_j \geq c_j \geq 1$.
- *Ordering,* where the learner is instructed to order k_j statements in each of J items (j = 1, … , J).
- *Matching pairs,* where the learner is instructed to match k_j (part)statements with k_j matching (part)statements. (This can be analyzed as *Ordering*.)

Figure 3.
Illustrations of different techniques of assessments.



3.2. Scoring Competence in Multiple-Choice Tests

A *Multiple-choice* test is a common technique used for scoring competence based on assessments. Traditionally, multiple choice tests have been scored using a conventional *Number Right* scoring method. Correct answered items are scored with a value of 1, incorrect and omitted items with a value of zero, and the sum of the scores on the J items is the score of the assessment.

Let X_j (j = 1, … , J) be a stochastic variable characterizing the score for answering item *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 such a Multiple-choice test the expected value of each X_j will be positive if a learner doesn't know the correct answer and is just guessing.

$$E(X_j) = 1 \cdot \frac{1}{k_j} + 0 \cdot \frac{k_j - 1}{k_j} = \frac{1}{k_j} > 0.$$

Therefore, 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. Various scoring methods have been presented to correct for this bias.

The *Rights Minus Wrongs* (or *Negative Marking*) scoring method penalizes the learner for incorrect responses. The fundamental idea behind this 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. If the learner decides to guess with such a scoring method, the expected score is

$$E(X_j) = 1 \cdot \frac{1}{k_j} + (-1) \cdot \frac{k_j - 1}{k_j} = -\frac{k_j - 2}{k_j} \leq 0.$$

The more distractors, the less the expected score becomes, 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 is zero if a learner is guessing. For this to happen when the reward for getting the question correct is 1, the penalty for an incorrect answer must be *-1/(k_j-1)*. That is,

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

Now, in a Multiple-choice test with k_j alternatives the expected score on each item is

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

This kind of neutrality with respect to guessing may be beneficial because of the not too strict discouragement to guess.

3.3. Partial Knowledge in Multiple-Choice Tests

Learners may be able to determine that some of the statements are incorrect even though they cannot identify the correct one, though. Being able to eliminate some such distractors is reflecting what is called "*partial competence*" and "*partial knowledge*".

If a learner eliminates *one* distractor correctly in a Multiple-choice test, the expected score is

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

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

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

And so on.

Table 1.
Expected scores after correctly elimination of distractors.

$E(X_j)$		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, as a learner who is not able to identify any distractors (i.e., an incompetent 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.

In line with this, we define a learner

- to have *full knowledge* if $E(X_j) = 1$
- to have *partial knowledge* if $0 < E(X_j) < 1$
- to be *incompetent (non-competent)* if $E(X_j) \leq 0$.

A variety of different partial-credit scoring methods are presented in the literature (Bush, 2001, Frary, 1989, Kurz, 1999, Traub et al., 1969)

3.4. Desirable Characteristics

Based on this discussion for the Multiple-choice technique, we suggest the following preliminary characteristics of a favorable scoring method. Let X_j = Score on item j, and

1. Completely correct $\Rightarrow X_j = 1$
2. Result as good as expected when guessing $\Rightarrow X_j = 0$
3. Result worse than expected when guessing $\Rightarrow X_j < 0$
4. Neutrality with respect to guessing $\Rightarrow E(X_j) = 0$
5. Partial knowledge is rewarded $\Rightarrow E(X_j \mid \text{Partial knowledge}) > 0$

We will in the following chapters discuss what these requirements implies with different scoring techniques and expand the list to identify the score to reward partly correctly answered items (indicating partial knowledge).

4. SCORING COMPETENCE IN MULTIPLE RESPONSE TESTS

4.1. Scoring Competence in Multiple Response Tests with Known c_j

4.1.1. Rewarding Only Completely Correct Responses

Let the learner be instructed to identify the c_j ($1 < c_j < k_j$) correct statements out of k_j statements in each of J items. Let the sum of the scores on the items be the score of the assessment, $X = \sum_{j=1}^{J} X_j$. Assume completely correct items will be scored with a value of 1, and incorrect and omitted answers with a value of zero. That is,

$$X_j = \begin{cases} 1 & if\ completely\ correct \\ 0 & if\ not\ completely\ correct \\ 0 & if\ omitted \end{cases}.$$

In such a multiple response test the expected value of each score will be positive if a learner is informed about the number of correct statements, is suggesting c_j statements as correct, but doesn't know which statements are correct, and is guessing.

$$E(X_j) = 1 \cdot \frac{1}{\binom{k_j}{c_j}} + 0 \cdot \frac{\binom{k_j}{c_j}-1}{\binom{k_j}{c_j}} = \frac{1}{\binom{k_j}{c_j}} > 0.$$

Introducing the discouraging factor $-\frac{1}{\binom{k_j}{c_j}-1}$ as a penalty for a not completely correct answer, then

$$X_j = \begin{cases} 1 & if\ completely\ correct \\ \dfrac{-1}{\binom{k_j}{c_j}-1} & if\ not\ completely\ correct \\ 0 & if\ omitted \end{cases}$$

and

$$E(X_j) = 1 \cdot \frac{1}{\binom{k_j}{c_j}} + \left(-\frac{1}{\binom{k_j}{c_j}-1}\right) \cdot \frac{\binom{k_j}{c_j}-1}{\binom{k_j}{c_j}} = 0.$$

With a given number of correct alternatives (c_j) the discouraging factor is decreasing with increasing number of statements (k_j). This seems "fair" as the item then becomes more difficult to answer completely correct when guessing. The effect of guessing is neutral, but if a learner manages to correctly identify some statements (either as correct or as incorrect) before guessing, guessing will have a positive expected effect reflecting "*partial knowledge*".

4.1.2. Rewarding Partial Knowledge

It seems reasonable also to reward a learner for answering partial correct on an item. If the scoring rule only is taking into account the number of correctly identified correct statements, though, the optimal strategy for the learner would be always to suggest that all k_j statements are correct. Instead, let's introduce,

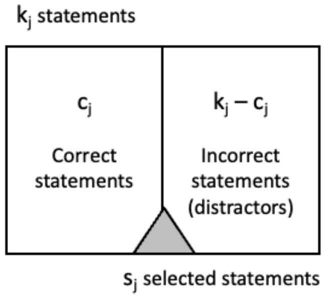
$$X_j = \frac{Y_j}{k_j} = \frac{\#correctly\ identified\ correct\ statements + \#correctly\ identified\ incorrect\ statements}{k_j}.$$

It seems too spendable to reward a learner for managing to identify less than the expected correctly identified statements if she is just guessing. Therefore, we suggest the scoring

- let $X_j = Y_j/k_j$ when $Y_j > E(Y_j)$
- let $X_j = 0$ when $Y_j = E(Y_j)$
- introduce a penalty such that the overall expected value of the score will be neutral, i.e. $E(X_j) = 0$, when $Y_j < E(Y_j)$
- let $X_j = 0$ if omitted.

Observe that given the learner is guessing and is proposing $s_j = c_j$ statements as correct, the number of correctly identified correct statements, V_j, will follow a Hypergeometric probability distribution, $V_j \sim Hyp(k_j, c_j, s_j)$.

Figure 4.
Illustration of a Hypergeometric experiment.



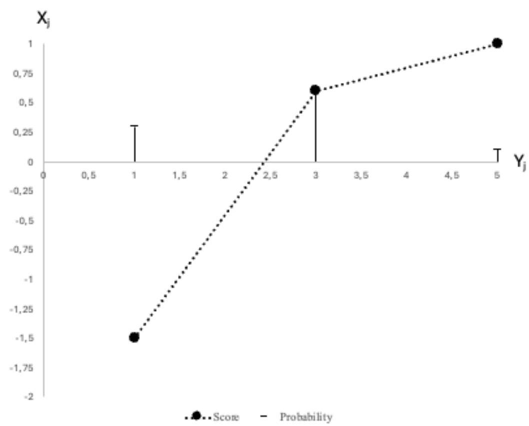
Example

Let $k_j = 5$ statements of which $c_j = 3$ are correct, $k_j - c_j = 5 - 3 = 2$ are distractors, and the learner is suggesting $s_j = c_j = 3$ statements as correct. (See figure 4.) The number of correctly identified correct statements, V_j, will follow a Hypergeometric probability distribution Hyp(k_j=5, c_j=3, s_j=c_j=3) and

$$V_j = 1 \Leftrightarrow Y_j = 1 \text{ and } P(Y_j = 1) = P(V_j = 1) = 0.3$$
$$V_j = 2 \Leftrightarrow Y_j = 3 \text{ and } P(Y_j = 3) = P(V_j = 2) = 0.6$$
$$V_j = 3 \Leftrightarrow Y_j = 5 \text{ and } P(Y_j = 5) = P(V_j = 3) = 0.1$$

If a learner is guessing, the expected number of correctly identified statements (both correct and incorrect) will be $E(Y_j) = 2.6$. Therefore, as the outcomes $Y_j = 3$ and $Y_j = 5$ are better than the expected value when guessing they are given scores 3/5 and 5/5 = 1, respectively. Only the outcome $Y_j = 1$ is worse than the expected value when guessing and to ensure a neutrality with respect to guessing the penalty for $Y_j = 1$ will be (approx.) -1.5.

Figure 5.
Scoring based on Y_j.



Observe that a "good" strategy for an incompetent learner, instead of guessing, in this situation would be suggesting that all statements are correct. As $Y_j = 3 > E(Y_j) = 2.6$, this would give a score 3/5. With multiple response tests we therefore suggest letting the number of correct statements be the result of a stochastic experiment and unknown to the learner, in any case when the relative number of correct statements is "large".

Example

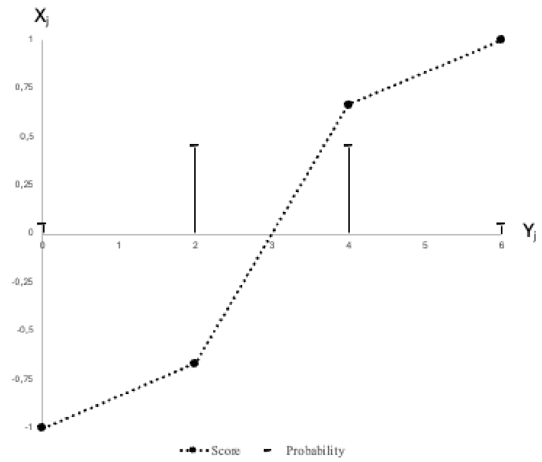
Let there be $k_j = 6$ statements of which $c_j = 3$ are correct, $k_j – c_j = 6 – 3 = 3$ are distractors, and the learner is suggesting $s_j = c_j = 3$ statements as correct. (See figure 4.)

The number of correctly identified correct statements, V_j, will follow a Hypergeometric probability distribution Hyp(k_j=6, c_j=3, s_j=c_j=3) and

$$V_j = 0 \Leftrightarrow Y_j = 0 \text{ and } P(Y_j = 0) = P(V_j = 0) = 0.05$$
$$V_j = 1 \Leftrightarrow Y_j = 2 \text{ and } P(Y_j = 2) = P(V_j = 1) = 0.45$$
$$V_j = 2 \Leftrightarrow Y_j = 4 \text{ and } P(Y_j = 4) = P(V_j = 2) = 0.45$$
$$V_j = 3 \Leftrightarrow Y_j = 6 \text{ and } P(Y_j = 6) = P(V_j = 3) = 0.05$$

If a learner is guessing, the expected number of correctly identified statements (both correct and incorrect) will be $E(Y_j) = 3$. Therefore, as the outcomes $Y_j = 4$ and $Y_j = 6$ are better than the expected value when guessing they are given scores 4/6 and 6/6=1, respectively. The outcomes $Y_j = 0$ and $Y_j = 2$ is worse than the expected value when guessing. To ensure a neutrality with respect to guessing there is an infinite number of possible values for the penalties. We suggest setting the penalties to ensure the kind of symmetry suggested in figure 2c. For $Y_j = 0$ we suggest the score $X_j = -1$, and for $Y_j = 2$ we suggest the score $X_j = -2/3$.

Figure 6.
Scoring based on Y_j.



4.2. Scoring Competence in Multiple Response Tests with Unknown C_j

4.2.1. Rewarding Partial Knowledge

Let the learner be instructed to identify the C_j correct statements out of k_j statements in each of J items. Let the sum of the scores on the items be the score of the assessment, $X = \sum_{j=1}^{J} X_j$. Assume the test designer is deciding to include the number of correct statements, $C_j = c_j$, according to a discrete probability distribution, and that the positions of these c_j correct statements are randomly selected among the k_j positions.

We recommend utilizing random generators both when deciding how many correct statements (c_j out of k_j) should be presented to the learner, and in which order the different statements should be presented. If not, it's easy presenting correct statements first and the distractors later. (It is when the test designer must be creative and invent the distractors that more thinking is required, thus making it tempting to postpone this task.) Our experience is that learners are quick to identify any such bias (even when it's not present!), which will affect the validity of the test. Furthermore, it is good pedagogy to minimize the number of distractors.

Example

Assume

- k_j = Total number of statements = 5
- $C_j = c_j = 3$ The number of correct statements as an outcome of a discrete probability distribution (for example a uniform U[1, 2, 3, 4, 5])
- $S_j = s_j$ The number of suggested correct statements by the learner, $S_j \sim U[0, 1, 2, 3, 4, 5]$
- C_j and S_j are independent stochastic variables
- V_j = The number of correctly identified correct statements
- Y_j = The sum of the number of correctly identified correct statements and the number of correctly identified incorrect statement.

Then the possible outcomes on item j will be as presented in table 2.

Table 2.
Possible outcomes on an item.

S_j	0	1	2	3	4	5
V_j	0	0, 1	0, 1, 2	1, 2, 3	2, 3	3
Y_j	2	1, 3	0, 2, 4	1, 3, 5	2, 4	3

As

$$P(Y_j = y_j) = \sum_l \mathrm{P}(S_j = s_l, V_j = v_l) = \sum_l \mathrm{P}(S_j = s_l)\, \mathrm{P}(V_j = v_l \mid S_j = s_l)$$

we see from table 2 and utilizing that $V_j \mid S_j$ is Hypergeometric distributed that

$$\begin{aligned} P(Y_j = 4) &= P(S_j = 2)\, P(V_j = 2 \mid S_j = 2) + P(S_j = 4)\, P(V_j = 3 \mid S_j = 4) \\ &= 1/6 \cdot 0.3 + 1/6 \cdot 0.4 \approx 0.12 \end{aligned}$$

The probability distribution of Y_j will be (approx.) as presented in table 3.

Table 3.
The probability distribution of Y_j.

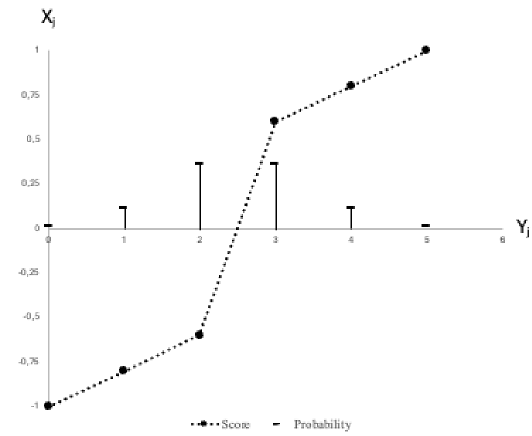
y_j	0	1	2	3	4	5
$P(Y_j = y_j)$	0.02	0.12	0.37	0.37	0.12	0.02

As the expected value of Y_j, $E(Y_j) = 2.5$, we suggest scoring the larger than expected values according to the same principle as suggested earlier (see chapter 3.1.2). That is,

$$Y_j = 3 \Rightarrow X_j = 3/5$$
$$Y_j = 4 \Rightarrow X_j = 4/5$$
$$Y_j = 5 \Rightarrow X_j = 5/5 = 1$$

To ensure the neutrality with respect to guessing there is an infinite number of possible values for the penalties. Again, we suggest setting the penalties to ensure the kind of symmetry suggested in figure 2c. For $Y_j = 0$ we suggest the penalty $X_j = -1$, for $Y_j = 1$ we suggest $X_j = -4/5$, and for $Y_j = 2$ we suggest $X_j = -3/5$.

Figure 7.
Scoring based on Y_j.



This procedure is also rewarding partial competence.

5. SCORING COMPETENCE IN ORDERING TESTS

5.1. Rewarding Only Completely Correct Responses

Let the learner be instructed to order k_j statements in J items in a specific order. This has a parallel to the famous "*matching hats problem*" within combinatorics originally described by Montmort in 1713 (Hathout, 2003). When scoring such an ordering test with k_j elements, a scoring procedure with penalty that effectively discourages guessing, and keeps a neutrality with respect to guessing, may be defined by

$$X_j = \begin{cases} 1 & if\ completely\ correct \\ -\frac{1}{k_j!-1} & if\ not\ completely\ correct. \\ 0 & if\ omitted \end{cases}$$

The neutrality is established as the expected score on each item is

$$E(X_j) = 1 \cdot \frac{1}{k_j!} + (-\frac{1}{k_j!-1}) \cdot \frac{k_j!-1}{k_j!} = 0.$$

With this scoring method, the effect of guessing is neutral, but if a learner manages to correctly eliminate some alternatives guessing will again give a positive expected effect. This is suggesting that with some partial knowledge it will be beneficial to guess.

5.2. Matching Hats Problem

5.2.1. Matching Hats Problem. Part 1

A group of k men enter a restaurant and check their hats. The hat-checker is absent minded and upon leaving she redistributes the hats back to the men at random.

Introduce Z_i = The i-th man leaves with his own hat (i = 1, 2, … , k.). Then,

$$P(Man\ i\ leaves\ with\ his\ own\ hat) = P(Z_i) = \frac{(k-1)!}{k!} = \frac{1}{k}$$

Similarly,

$$P(Man\ i\ and\ j\ leaves\ with\ their\ own\ hats) = P\left(Z_i \cap Z_j\right) = \frac{(k-2)!}{k!}$$

$$\mathrm{P}(Man\ i, j\ and\ l\ leaves\ with\ their\ own\ hats) = P\left(Z_i \cap Z_j \cap Z_l\right) = \frac{(k-3)!}{k!}$$

and so on. Now,

P(At least one man leaves with his own hat)

$$= P(\cup_{i=1}^{k} Z_i)$$

$$= \sum_{i=1}^{k} P(Z_i) - \sum_{i<j} P(Z_i \cap Z_j) + \cdots + (-1)^{k+1} P(Z_1 \cap \ldots \cap Z_k)$$

$$= \binom{k}{1}\frac{(k-1)!}{k!} - \binom{k}{2}\frac{(k-2)!}{k!} + \binom{k}{3}\frac{(k-3)!}{k!} + \cdots + (-1)^{k+1}\binom{k}{k}\frac{(k-k)!}{k!}$$

$$= \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} \ldots + (-1)^{k+1}\frac{1}{k!}$$

Furthermore,

P(No man leaves with his own hat)

= 1 - P(At least one man leaves with his own hat)

$$= 1 - [\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} \ldots + (-1)^{k+1}\frac{1}{k!}]$$

$$= \frac{D(k)}{k!}$$

where $D(k)$ is the number of *derangements* of the k items (i.e., a permutation of the elements of a set in which no element appears in its original position). Consequently,

$$D(k) = k!\ \{1 - [\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} \ldots + (-1)^{k+1}\frac{1}{k!}]\}$$

Introduce

Y = # correctly returned hats (out of k).

Then,

$$P(Exactly\ y\ correctly\ returned\ hats\ out\ of\ k) = P(Y = y) = \frac{D(k-y)}{y!\,(k-y)!}$$

5.2.2. Matching Hats Problem. Part 2

Observe that

D(1) = 0

D(2) = 1

D(k) = (k-1){D(k-1) + D(k-2)}

introduces an alternative, and maybe more convenient, way of calculating D(k).

5.2.3. Matching Hats Problem. Part 3

Observe that

$$E(Y) = \sum_{y=0}^{k} y\, P(Y = y) = 1.$$

That is, *we always expect exactly one match*, regardless of how many men are present. This can be deduced as follows. Let

$$Y = Z_1 + Z_2 + \ldots + Z_k$$

where Z_i is 1 if the i-th man gets his own hat back, and 0 otherwise. Since

$$E(Y) = E(Z_1) + E(Z_2) + ... + E(Z_k),$$

and

$$E(Z_i) = 0 \cdot P(Z_i = 0) + 1 \cdot P(Z_i = 1) = P(Z_i = 1) = 1/k \text{ for } i = 1, 2, 3, ... , k$$

it follows that

$$E(Y) = 1/k + 1/k + ... + 1/k = 1.$$

Thus, no matter how many men are involved, we expect that exactly one man receives his own hat.

5.3. Rewarding Partial Knowledge

Let

Y_j = # correctly placed statements (out of k_j) in item j (j = 1, … , J).

Then, the probability distribution of Y_j is (see subchapter 4.2.1)

$$P(Y_j = y) = \frac{\sum_{i=0}^{k_j - y} \frac{(-1)^i}{i!}}{y!} \text{ ; } y = 0, 1, \dots, k_j$$

Table 4.
The probability distribution of Y_j for selected values of k_j.

P(y)	**k_j**						
y	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{3}{8}$	$\frac{11}{30}$	$\frac{53}{144}$	$\frac{309}{840}$	$\frac{2119}{5760}$
1		$\frac{1}{2}$	$\frac{1}{3}$	$\frac{3}{8}$	$\frac{11}{30}$	$\frac{53}{144}$	$\frac{103}{280}$
2	$\frac{1}{2}$		$\frac{1}{4}$	$\frac{1}{6}$	$\frac{3}{16}$	$\frac{11}{60}$	$\frac{53}{288}$
3		$\frac{1}{6}$		$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{16}$	$\frac{11}{180}$
4			$\frac{1}{24}$		$\frac{1}{48}$	$\frac{1}{72}$	$\frac{1}{64}$
5				$\frac{1}{120}$		$\frac{1}{240}$	$\frac{1}{360} = .003$
6					$\frac{1}{720}$		$\frac{1}{1440}$
7						$\frac{1}{5040}$	
8							$\frac{1}{40320}$

Remember that $E(Y_j) = 1$, independent of the value of k_j. That is, when a learner is guessing, we're always expecting exactly one correctly placed statement. Therefore, we feel a fair score for exactly one correctly placed statement should be 0. (See row 2 in table 5 below.) We suggest a score of 1 for all statements placed in the correct order.

That is, the score for $Y_j = 1$ and $Y_j = k_j$ are given, and we need to decide on the scores for the values y = 0, 2, 3, …, k_j-2. (Remember y = k_j-1 is impossible.) We find it reasonable to suggest,

$$\text{score for exactly } y_j \text{ correctly placed statements out of } k_j,\ X_j = \frac{y_j}{k_j}$$

for $y_j > 1$ and $k_j > 4$. (For the special cases of k_j = 2, 3, 4, see our suggestions in table 5.)

Furthermore, we suggest introducing a penalty for no correctly placed statements (i.e., a result worse than what is expected when guessing), such that the expected score will become 0. Then, for some selected values of k_j, the scoring will be as presented in table 5 below.

Table 5.
Suggested scoring resulting in neutrality when guessing.

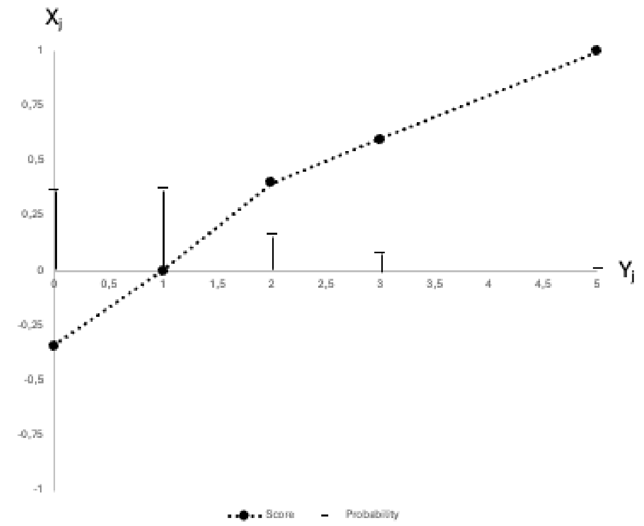
Score	**k_j**						
y_j	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	-1	$-\frac{1}{2}$	$-\frac{4}{9}$	$-\frac{15}{44}$	$-\frac{228}{795}$	$-\frac{9555}{38934}$	$-\frac{3186}{14833}$
1		0	0	0	0	0	0
2	1		$\frac{1}{2}$	$\frac{2}{5}$	$\frac{2}{6}$	$\frac{2}{7}$	$\frac{2}{8}$
3		1		$\frac{3}{5}$	$\frac{3}{6}$	$\frac{3}{7}$	$\frac{3}{8}$
4			1		$\frac{4}{6}$	$\frac{4}{7}$	$\frac{4}{8}$
5				1		$\frac{5}{7}$	$\frac{5}{8}$
6					1		$\frac{6}{8}$
7						1	
8							1

We see that the penalty for having no correctly placed statements is reduced with increasing difficulty (i.e., with increasing number of alternatives). The effect of guessing is neutral, but if a learner manages to correctly eliminate some alternatives before guessing, guessing will have a positive expected effect reflecting "*partial knowledge*".

Example

Let k_j = 5. With data from table 5, the scoring may be illustrated as in figure 8.

Figure 8.
Scoring based on Y_j.



6. CONCLUSION

Based on the discussion above, we suggest using either Multiple-choice tests with a random number of correct alternatives, where the number is not communicated to the learner, or some kind of ordering test to estimate a learner's competence. We suggest giving the score 1 for a completely correct answer, a positive score (depending on the test) for partial correct answers better than expected if the learner is guessing, 0 for a result exactly as expected if guessing, and introducing penalties for answers not as good as expected if guessing. The size of the penalty should ensure that the expected score, when guessing, is 0. All suggested techniques of scoring are taking into account partial knowledge.

REFERENCES

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

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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press.

Hathout, H. (2003). The old hats problem. *Rose-Hulman Undergraduate Mathematics Journal*, *4*(1), Article 2. https://scholar.rose-hulman.edu/rhumj/vol4/iss1/2

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

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

Salmon, P. M., Stanton, N. A., Walker, G. H., & Jenkins, D. P. (2009). *Distributed Situation Awareness: Theory, Measurement and Application to Teamwork*. Surrey: Ashgate.

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), 847-861. doi: 10.1177/001316446902900410

AUTHORS' INFORMATION

Full name: Magne Vollan Aarset
Institutional affiliation: Norwegian university of science and technology - NTNU
Institutional address: NTNU, Box 1517, NO-6025 Ålesund, Norway
E-mail address: magne.aarset@ntnu.no, magneaarset@gmail.com
Short biographical sketch: Magne V. Aarset currently works at the Norwegian University of Science and Technology – NTNU and as CEO at Statcon AS. His formal education is within Mathematics, Mathematical Statistics, Psychology and Music. He has 20 years working experience mainly within Aviation and Aerospace, and 20 years within the University sector. His research is within Cognitive Psychology (with special interest in distributional situational awareness and distributed cognition), Risk Management (including heuristics and decision analysis), and Mathematical Statistics (Machine Learning – AI and Bathtub Failure Rates) often applied within Leadership or within Learning.

Full name: Stig Ytterstad
Institutional affiliation: BI Norwegian Business School
Institutional address: Nydalsveien 37, NO-0484 Oslo, Norway
E-mail address: stig.ytterstad@bi.no
Short biographical sketch: Stig Ytterstad is Teaching Professor Emeritus at BI The Norwegian Business School. His formal education is within History, Political Science, Law, Pedagogic, Organization Psychology and Leadership. He has 16 years working experience in the school system, mainly at high school, and 31 years within the University sector. His research is within Transformational Leadership (with special interest in learning) and Change Management and Learning Styles (with focus on Sternbergs Thinking Styles).