RESEARCH NOTE

NOTA DE INVESTIGACIÓN

RESEARCH NOTE/NOTA DE INVESTIGACIÓN

Students as Rational Actors in Multiple-choice Tests and How to Mark their Mistakes

El estudiante como actor racional ante exámenes tipo test y cómo puntuar sus errores

Daniel Guinea-Martin
University of Malaga, Spain
daniel.guinea@uma.es

Received/Recibido: 10/06/2023
Accepted/Aceptado: 25/07/2023

ABSTRACT

Sociologists teaching at universities have seen how multiple-choice tests with closed-ended questions have become the norm in the discipline. If wrong answers are not penalised, these tests become a context that favours the emergence of rational actors in the form of the student-player. Such students minimise their effort by taking advantage of the probability of answering correctly out of luck. By applying the definition of a Bernoulli random variable, this methodological note presents the score that must be awarded to incorrect answers in exams, regardless of the number of questions and response options. Deviating from this score means one of two things: favouring the emergence of the student-player, or overly penalising the risks taken by students when sitting exams.

KEYWORDS: closed-ended question; evaluation; exam; mark, penalty; rational action; score; student; test.

HOW TO REFERENCE: Guinea-Martín, D. (2023). El estudiante como actor racional ante exámenes tipo test y cómo puntuar sus errores. *Revista Centra de Ciencias Sociales*, 2(2), 97-114. https://doi.org/10.54790/rccs.69

The Spanish (original) version can be read at https://doi.org/10.54790/rccs.69

RESUMEN

La experiencia de los sociólogos que enseñamos la disciplina en la universidad muestra que el uso de los exámenes tipo test con respuestas cerradas se ha generalizado en nuestra disciplina, entre otras. Si no se penalizan las respuestas erróneas, estos test se transforman en un contexto que favorece el surgimiento de un actor racional en la forma del estudiante-jugador. Este minimiza su esfuerzo aprovechando la probabilidad de acertar con que le obsequia la suerte. Aplicando la definición de variable aleatoria de tipo Bernoulli, en esta nota metodológica presentamos la puntuación que deben tener las respuestas erróneas en exámenes de cualquier extensión en su número de preguntas y opciones de respuesta. Desviarse de esta puntuación supone, o bien favorecer el surgimiento del estudiante-jugador, o bien penalizar en exceso la asunción del riesgo a equivocarse que todo estudiante debe afrontar en un test.

PALABRAS CLAVE: acción racional; estudiante; evaluación; examen; nota; penalización; pregunta cerrada; puntuación; test.

1. Introduction

Multiple-choice exams (referred to hereinafter as "exams" or "tests") have been the subject of sociological reflection for more than fifty years (see, for example, Goslin and Glass, 1967). Nevertheless, their expansion beyond controversial intelligence tests and American university entrance exams is more recent. As Edwards (2006) points out, their proliferation can be linked to Parsons' claims about school: in meritocratic societies, this institution fulfils the dual function of (a) talent selection, and (b) the subsequent allocation of individuals to jobs (see, however, Bennett de Marrais and LeCompte, 1998, for a critique of this argument).

In recent years, these tests have been taken in numerous Spanish university degrees, including the social sciences, where discursive subjects abound and, therefore, it does not appear to be the best way to evaluate the knowledge acquired. It is the task of sociologists to study the causes of this trend. Although this is not the aim of this brief methodological note, it should be mentioned that the most cited reason, in my experience, is that students and teachers prefer tests because they require less effort than exams with open-ended questions: not only do teachers correct them more easily and quickly, but the conflict in mark reviews or appeals is also reduced. Furthermore, students are spared having to write their own rationale, as well as the prospect of making grammatical and spelling errors when answering the questions. Another advantage of tests is that, even if the decisions regarding their design and contents are subjective, the standardisation of their closed answers makes their evaluation more objective (Andreasen, Rasmussen and Ydesen, 2013).

Nonetheless, the objective of this methodological note is to point out that those tests in which wrong answers are not penalised is a social context that favours the rational calculation of the well-known *homo economicus* (Levitt and List, 2008). In this case, the ideal typical model, in the Weberian sense (Weber, [1922] 1982), is a student solely motivated by the aim of passing with the least effort (i.e., without studying). This stereotypical idle student answers the test questions at random. Faced with this situation, many teachers are unaware of how wrong answers should be evaluated to reduce the student-player's expectation of passing to zero. In some cases there are no penalties, thereby giving a generous advantage to the student who has not studied. As will be seen later, this is equivalent to the probability of randomly guessing the correct answer from each question's response options. (Several studies, for example, Dehnad, Nasser and Hosseini, 2014, and Schneid et al., 2014, conclude that three response options is the optimal number to maximise the validity and reliability of the tests). Translated into points, the student-player would obtain on average the maximum exam score, for example, 10, multiplied by the previous probability; for example, if there are 3 possible answers, on average the student's final score would be $10 \cdot \frac{1}{3} = 3.3$ points.

In other cases they penalise too little, which continues to give an advantage to the idle student which will be within the range $\left(0, \frac{1}{\text{number of possible answers}}\right)$, depending on the penalty used.

Finally, in some instances, incorrect answers are over-penalised in relation to that which would eliminate the role of chance. This situation shifts the penalty to the left on the real line: that of the student-player becomes negative, while that of the student who has prepared becomes lower than the numerical value that best reflects their knowledge.

Although teachers sometimes miscalculate how to penalise incorrect answers, some students employ strategies designed to score as high as possible on tests with as little study as possible (see, for example, Psiconociendo, 2022, and Sentipia, 2022). There is even a campaign for the Spanish Constitutional Court to declare the penalty for incorrect answers on tests to be unlawful (see Icaro100, 2010) based on the legal justification proposed by Muñoz Clares and Caballero Salinas (2019).

In light of this situation, this methodological note is intended to help both teachers and students approach the issue (a) knowing its technical details, and (b) being aware of the model student who is favoured by one kind of score or another, no penalty vs. a fair penalty: the student-player vs. the responsible student, respectively. The final result of a correct calibration of the score, correcting according to wrong answers, is that the final exam mark numerically reflects the student's state of knowledge.

Furthermore, and unlike some articles that deal with the same subject and that have served as inspiration for this note (see, for example, Morales, 2017; among the many resources on these topics that can be found online, of particular note are those by the American Statistical Association, 2013; Bickis, 2017 and Stanbrough, 2009), the general formula is presented here to calculate the mark for wrong answers, regardless of the number of questions in the exam and the number of response options for each. The only constant assumption is that there is a total of ten marks available on the exam, as is usual in the Spanish school and university system. (In any case, if another score were used, it would suffice to replace the number 10 with the corresponding number in the formulae below).

The result is a simple, but not trivial calculation. The calculation requires the command of some basic statistical concepts: sample space, Bernoulli and binomial discrete random variables, as well as its expected value and probability density function (pdf) also known as probability mass function (pmf). Those concepts, that are discussed more extensively in any statistical manual, will be explained briefly and exemplified in this note (see, for example, Martín-Pliego and Ruiz-Maya, 2006).

2. The rational actor and the role of chance in tests

If the model of the student who is motivated exclusively by the aim of passing with the least effort is taken to extremes, this can be equated to not studying at all. However, even without studying, if there is no penalty for an incorrect answer, such a student *qua* rational agent answers the closed-ended questions because they can guess the correct answer from the response options by pure chance.

Randomly choosing an answer out of the *m* options has a probability $p = \frac{1}{m}$ of being right, and a probability $q = 1 - p = 1 - \frac{1}{m} = \frac{m-1}{m}$ of being wrong. Indeed, the probability of randomly getting all the exam questions right is lower, since, if each question is regarded as independent to the other questions, which is reasonable for the case of the student who has not studied anything. Then the probability of getting the *n* exam questions right is $p^n < p$.

This situation replicates the games of chance created by the Swiss mathematician Jacob Bernoulli ([1713] 1993) in the eighteenth century, thus becoming known as Bernoulli's experiments or trials. The fundamental issue is this: those students who do not study and who choose their answers at random are rewarded if an exam's marking criteria award positive points for correct answers, for example, one point, and zero points for both unanswered questions and incorrect answers. In this case, the student-player can expect to get $1 \cdot p + 0 \cdot q = 1 \cdot \frac{1}{m} + 0 \cdot \frac{m-1}{m} = \frac{1}{m} = p$ points on each exam question, a score that is by no means negligible.

The aim of this article is to explain the above calculation, as well as the marking criteria that would entirely cancel out such an a priori advantage, so that, on average, the student-player gets a zero on the exam.

3. Fundamental concepts

Answering closed-ended questions is an example of a "Bernoulli" trial or experiment because there are only two possible outcomes for the student: success or failure, which can be denoted ω_1 and ω_2, respectively, and which make up the so-called "sample space" Ω of the test. However, it is not possible to operate mathematically with events. Therefore, the "random variable" function is introduced (hereinafter, "RV"); in this case, $Y \equiv$ "randomly guessing the answer to a question", which translates, so to speak, the two events $\Omega = \{\omega_1, \omega_2\}$ from the sample space Ω to the two numerical values of the RV. Y: $Y = \{y_1, y_2\}$. For this reason, it is common to see the following notation in statistical manuals:

$$(1)\ Y : \Omega \rightarrow R.$$

The expression (1) means that the RV Y is a function defined on the sample space (set of the results of a random experiment) that takes values in the body of the real numbers R. However, in Bernoulli's experiments or trials, the RV function translates the test results to the field of Z integers:

$$Y{:}\Omega \rightarrow Z$$

Specifically, as done previously, the integer 1 is usually reserved (for the sake of clarity and operational convenience) for the "success" of the test (answering correctly in the example given here corresponds to the value 1 of the RV: $Y = y_1 = 1$) and the integer zero as the opposite, for the "failure" or mistake ($Y = y_2 = 0$) (Baclawski, 2008, p. 48).

Each value y_1, y_2 of the RV Y has an associated probability of being verified in a given trial or experiment or, in the context of this article, in each exam question: the student answers correctly ($\omega_1 = \text{success}, y_1 = 1$) with a probability $p = \frac{1}{m}$ and the student answers incorrectly ($\omega_1 = \textit{failure}, y_2 = 0$) with a probability $\frac{m-1}{m}$. In summary, the RV has a probability distribution which, here, is also called Bernoulli's distribution and which is succinctly expressed as $Y \sim \text{Bernoulli}\ (p)$.

"Expectation" (a mathematical operator denoted by $E[\cdot]$) of an RV Y is the name given to its "expected value": $E[Y]$, which depends on the *pmf* and tends towards the Y values with the highest associated probability. When there are empirical data, the equivalent is the mean: this approaches the expectation as the sample increases or repeats[1].

In the discrete case, the expectation is calculated as the sum of the product of the probability (p_i) that each value i of the variable Y (y_i) has for that same value y_i (here, $i = \{1,2\}$): $E[Y] = \sum_i y_i\, p_i = y_1 p_1 + y_2 p_2 = y_1 \frac{1}{m} + y_2 \left(1 - \frac{1}{m}\right) = y_1 \frac{1}{m} + y_2 \frac{m-1}{m}$.

When the correct answer is marked with the integer 1 ($y_1 = 1$) and the incorrect answer with 0 ($y_2 = 0$), the expected value is p: $E[Y] = 1 \cdot \frac{1}{m} + 0 \cdot \frac{m-1}{m} = \frac{1}{m} = p$.

4. Extension to any type of score per question

Awarding the correct answer a score of 1 makes sense if: (a) the maximum mark that can be obtained is 10 points, representing a completely perfect answer, and (b) the exam consists of 10 questions of equal value; or if the exam has n questions and the final mark is not restricted to 10 points, but is $1 \cdot n = n$ points.

In the following discussion assumption (a) will be maintained, since the scale of marks from zero to ten points for evaluating exams is the most common in the Spanish educational system. However, assumption (b) is questionable as the number of questions asked varies in the tests; in fact, exams are rarely limited to only 10 questions, since they compensate for not having to reason explicitly by requiring answers to numerous questions.

The maximum exam score, 10, must be divided among the n questions, so that the value of the correct answer, u, must be $u = \frac{10}{n}$ points. For example, with n=20 questions, $u = \frac{10}{n} = \frac{10}{20} = \frac{1}{2} = 0.5$ points. This shows how the image or range of our Bernoulli RV (that is, the set of values it can take) corresponds to the set of rational numbers (Q) and not to that of the integers (Z). In other words, in these cases it is verified that:

$$Y : \Omega \rightarrow \mathrm{Q}$$

As mentioned previously, this particularity is not usually taken into account in presentations on how to mark multiple-choice exams (see, for example, Morales, 2017); however, it is paramount for resolving the crux of the matter: how to score each incorrect answer. What rational number should be assigned to y_2, the second possible value of the RV Y which represents the mistakes in the answers? This ignorance can be formally expressed in the *pmf* of the RV Y, that is, the function that informs the probability with which the RV Y adopts each of its y_i values, $P(Y{=}y_1)$ or $P(Y{=}y_2)$:

$$Y = \begin{cases} Y(\omega_1 = \text{success}) = y_1 = u = \frac{10}{n}, P\left(Y = y_1 = \frac{10}{n}\right) = \frac{1}{m} = p \\ Y(\omega_2 = \text{failure}) = y_2 = ?, P(Y = y_2 = ?) = \frac{m-1}{m} = q = 1 - p \end{cases}$$

A teacher who shares the value judgement that the mark of a student who has not studied must be the numerical equivalent of nothing, that is, a zero, will agree that the quantity y_2, for now unknown, must be such as to centre the expectation of the variable Y on zero, $E[Y]=0$, and not on $p = \frac{1}{m}$ as was the case with the Bernoulli RV Y with which this explanation began, that is, that which associates the zero value with mistakes. Therefore, and since the definition of mathematical expectation is $E[Y]=\sum_i y_i p_i$, the situation is represented as follows:

$$E[Y] = y_1 \cdot p + y_2 \cdot q = \frac{10}{n} \cdot \frac{1}{m} + y_2 \frac{(m-1)}{m}$$

For the expectation to be zero, $E[Y] = \frac{10}{n} \cdot \frac{1}{m} + y_2 \cdot \frac{m-1}{m} = 0$ the mark for a wrong answer, y_2, must adopt a precise value and, in order to determine it, the expression of the expectation equal to zero is written in terms of y_2:

$$\frac{10}{n}\cdot\frac{1}{m}+y_2\cdot\frac{m-1}{m}=0$$
$$y_2\cdot\frac{m-1}{m}=\frac{-10}{nm}$$
$$y_2=\frac{-10m}{nm(m-1)}=\frac{-10}{n(m-1)}=\frac{10}{n}\cdot\left(\frac{-1}{m-1}\right)=u\cdot\left(\frac{-1}{m-1}\right)=\frac{-u}{m-1}$$

(In the ensuing discussion it is sometimes convenient to simplify the notation by denoting with v the penalty $\frac{-1}{m-1}$ that is applied to each incorrect answer, such that $y_2=u\cdot\left(\frac{-1}{m-1}\right)=uv$.) In other words, in a multiple-choice exam with a maximum score of 10 points and n questions, where each question has m possible answers, by scoring each incorrect answer with $y_2=\frac{-10}{n(m-1)}=u\left(\frac{-1}{m-1}\right)=uv$, the student who guesses the answers randomly has an expected mark of zero.

Therefore, the *pmf* of the Bernoulli RV Y that takes values in the set of rational numbers Q and with expected value equal to 0:

$$Y=\begin{cases} Y(\text{success})=y_1=u=\frac{10}{n}, P(Y=y_1)=\frac{1}{m} \\ Y(\text{failure})=y_2=uv=u\cdot\left(\frac{-1}{m-1}\right)=uv, P(Y=y_2)=\frac{m-1}{m}\end{cases}$$

where we have the following: (a) the values $Y(\text{success})=y_1=u$ and $Y(\text{failure})=y_2=uv$; (b) n questions in the exam; (c) m possible answers per question.

In short, $\frac{10}{n}\cdot\left(\frac{-1}{m-1}\right)$ is the score that a wrong answer must have on a test so that, whoever randomly guesses the answer, obtains a zero on average. Note that this expression is a function of two parameters: the n questions that make up the exam and the m response options to each question. As mentioned previously, in this formulation the maximum score on the exam is 10 points. If this were not the case, there would be a third parameter k on which the correct and incorrect answer score would depend. This formula is reproduced in the conclusion in plainer language so that it transcends as the main message of this methodological note and interested teachers can easily apply it to their situation.

By way of verification, by making $y_2=uv$, the student-player does indeed get on average a zero in each question.

$$E[Y]=y_1\cdot p+y_2\cdot q=\frac{10}{n}\cdot\frac{1}{m}-\frac{10}{n(m-1)}\cdot\frac{(m-1)}{m}=\frac{10}{nm}-\frac{10}{nm}=0$$

Given this scoring scheme, the following facts are derived:

1. The maximum score on the exam with questions is effectively $n \cdot \frac{10}{n} = 10$;
2. the minimum is $n \cdot \frac{10}{n} \cdot \left(\frac{-1}{m-1}\right) = \frac{-10}{m-1}$;
3. If someone wants to adopt marking criteria that assign the value $w \neq uv$ to $Y(\text{wrong}) = y_2$, there are two possibilities:

   - if $w < uv$, the marking criteria penalise incorrect answers too severely, inhibiting the risk-taking involved in answering a question when there is the slightest doubt as to which is the correct option;
   - if $w > uv$, the marking criteria penalise incorrect answers too mildly, bringing the real student being examined closer to the ideal-typical model of the rational actor that has been dubbed here the "student-player".

4.1. Expected exam mark

By the linearity of the expectation operator, $E[\cdot]$, and assuming that the *n* questions give rise to *n* independent and identically distributed RV $Y_1, Y_2, \ldots, Y_n$ with the *pmf* described above, the expected overall exam mark will be equal to zero: $E[n \cdot Y] = n \cdot E[Y] = n \cdot 0 = 0$.

4.2. Examples

Let's start by posing an example where the pmf of the R.V. $Y \sim \text{Bernoulli}\left(p = \frac{1}{5}\right)$ reflects a test with $n = 30$ questions and $m = 5$ response options:

$$Y = \begin{cases} Y(\text{success}) = y_1 = \frac{10}{30} = \frac{1}{3}, P\left(Y = y_1 = \frac{1}{3}\right) = \frac{1}{5} \\ Y(\text{failure}) = y_2 = \frac{1}{3}\left(\frac{-1}{5-1}\right) = \frac{-1}{12}, P\left(Y = y_2 = \frac{-1}{12}\right) = \frac{4}{5} \end{cases}$$

Then it can be verified that:

$$E[Y] = \frac{1}{3} \cdot \frac{1}{5} - \frac{1}{12} \cdot \frac{4}{5} = \frac{1}{15} - \frac{4}{60} = \frac{1}{15} - \frac{1}{15} = 0$$

Another example is a test with $n = 20$ questions and $m = 3$ response options per question. Thus, each correct answer is worth $u = \frac{10}{20} = \frac{1}{2}$ points, and each incorrect answer should be given

$$\frac{1}{2}\cdot\left(\frac{-1}{3-1}\right)=\frac{-1}{2}\cdot\frac{1}{2}=\frac{-1}{4}=-0.25 \text{ points}$$

With this, the expectation of the RV Y equals zero:

$$E[Y]=y_1\cdot p+y_2\cdot q=\frac{1}{2}\cdot\frac{1}{3}-\frac{1}{4}\cdot\frac{2}{3}=\frac{1}{6}-\frac{1}{6}=0$$

If the same exam had questions with $m=4$ response options, each mistake should score $\frac{1}{2}\cdot\left(\frac{-1}{4-1}\right)=\frac{-1}{2}\cdot\frac{1}{3}=\frac{-1}{6}\approx-0.17$ points. Thus, in effect, the expectation is:

$$E[Y]=\frac{1}{2}\cdot\frac{1}{4}-\frac{1}{6}\cdot\frac{3}{4}=\frac{1}{8}-\frac{1}{8}=0$$

If, on the other hand, the exam had $n=40$ questions with $m=3$ response options, answering correctly would be worth $u=\frac{10}{40}=\frac{1}{4}$ points and answering incorrectly, $uv=\frac{1}{4}\cdot\left(\frac{-1}{3-1}\right)=\frac{-1}{4}\cdot\frac{1}{2}=\frac{-1}{8}=-0.125$ points. Thus, the expected value is:

$$E[Y]=\frac{1}{4}\cdot\frac{1}{3}-\frac{1}{4\cdot 2}\cdot\frac{2}{3}=\frac{1}{4}\cdot\frac{1}{3}-\frac{1}{4}\cdot\frac{1}{3}=0$$

Likewise, $m=4$ response options would also result in awarding an incorrect answer $\frac{1}{4}\cdot\left(\frac{-1}{4-1}\right)=\frac{-1}{4}\cdot\frac{1}{3}=\frac{-1}{12}\approx-0.083$ points. Indeed, in such a case the expectation remains at zero:

$$E[Y]=\frac{1}{4}\cdot\frac{1}{3}-\frac{1}{8}\cdot\frac{2}{3}=\frac{1}{12}-\frac{1}{12}=0$$

To summarise, the more response choices and questions there are in an exam, the lower the penalty for incorrect answers.

5. Number of correct answers in n questions: transition from the Bernoulli to the binomial distribution

On one occasion, a mistake was made in a step of the code used to randomly alternate the position of the correct answer in a test with $n=20$ questions and $m=3$ answer options. The result was that in the first exam sitting for the subject "Sociology of diversity", of the UNED Sociology degree, the correct answer was always the third response. A student who had not studied, and who resat the exam (second exam sitting), decided to continue opting for the answer (c) in each test question, when in that exam the correct answer varied randomly from question to question. Thus, the student randomly guessed the correct answer to 8 questions, and got 12 wrong, obtaining a score of $8 \cdot \frac{1}{2} - 12 \cdot \frac{1}{4} = 1$ "free" point, i.e., 10% of the total exam mark. However, if the incorrect answers had not been penalised as explained in the previous section, the student would have got $8 \cdot \frac{1}{2} = 4$ points by pure luck, 40% of the total mark. With this final result, the student could consider appealing the final mark for the course, arguing that it was very close to the exam pass mark, as is often the case among students whose mark is just below a five.

An interesting question is, what is the probability of obtaining 8 correct answers in 20 attempts? Obtaining *x* correct answers in *n* attempts or trials describes the data-generating process of a random variable with a binomial distribution. By way of example, this random variable is denoted $X \equiv$ "number of correct answers in n questions in a test". In addition, this statistical model assumes that each of the *n* test questions is independent of the others and that, in addition, the RV associated with each question is identically distributed, i.e., it follows a Bernoulli distribution with probability *p*[2].

Under these conditions, the probability of obtaining exactly *x* correct answers in *n* questions is expressed as $P(X = x)$, and the *pmf* of the variable $X \sim B(n, p)$ is:

$$P(X = x) = \binom{n}{x} p^x q^{n-x},\ x = 0,1, \dots ,20$$

Given that X is a binomial RV, its expectation is $E[X] = np$ (Martín-Pliego and Ruiz-Maya, 2006), thus, when $n = 20,\ p = \frac{1}{3}$, $E[X] = 20 \cdot \frac{1}{3} = 6.\overline{6}$. In the example cited, the result is somewhat higher than the expectation, since there were $x = 8$ correct answers. The probability of this occurring can be quantified as follows:

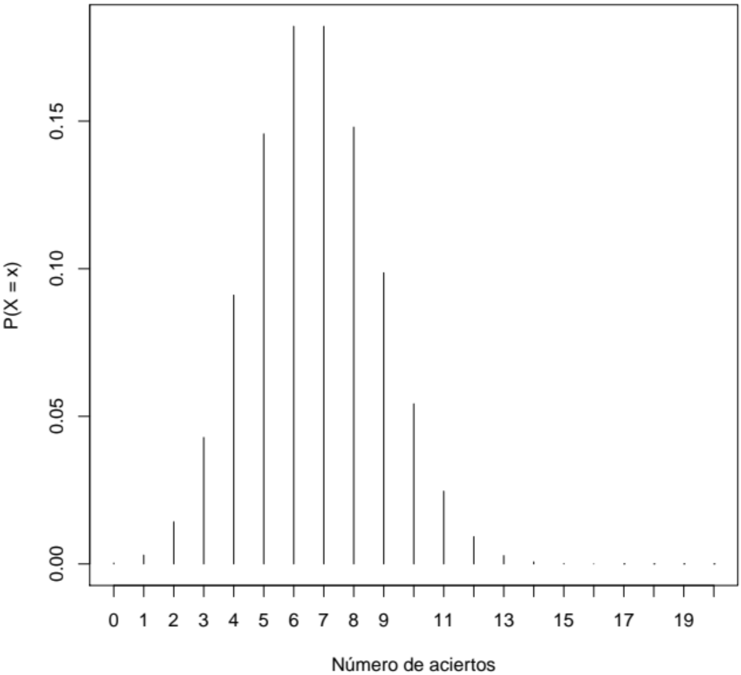
$$P(X = x = 8) = \binom{20}{8} p^8 q^{12} = \frac{20!}{12!\,8!}\, p^8 q^{12} = 125970 \cdot \left(\frac{1}{3}\right)^8 \cdot \left(\frac{2}{3}\right)^{12}$$
$$= 125970 \cdot 0{,}0001524158 \cdot 0{,}007707347 = 0{,}1479797.$$

In other words, around 15% of the student-players will get 8 correct by chance or, alternatively, if one of these students were to repeat the test *ad infinitum*, they would get 8 correct by chance almost 15% of the time.

What is the probability for the 20 possible outcomes, from zero to 20 correct answers obtained by chance? This is represented in Graph 1 of the *pmf* $X \sim B\left(n = 20, p = \frac{1}{3}\right)$:

Graph 1

Probability distribution of X∼B(n=20, p=1/3)



Graph 1 shows that the highest probability corresponds to the sample results that coincide with the integers around the mathematical expectation of $E[X]$: 6 correct answers and 14 incorrect answers, on the one hand, and 7 right and 13 wrong on the other. Indeed, this binomial distribution has two values with maximum probability, that is, two modes M_0, verifying the following inequality (the test is found in Arnáiz, 1986, cited in Martín-Pliego and Ruiz-Maya, 2006, p. 188):

$$\begin{aligned} np - q &\leq M_0 \leq np + p \\ 20 \cdot \frac{1}{3} - \frac{2}{3} &\leq M_o \leq 20 \cdot \frac{1}{3} + \frac{1}{3} \\ 6 &\leq M_o \leq 7. \end{aligned}$$

The exact value of this, the highest probability of this *pmf*, is:

$$P(X = 6) = \binom{20}{6} p^6 q^{14} = 0{,}1821288;$$

$$P(X = 7) = \binom{20}{7} p^7 q^{13} = 0{,}1821288.$$

In other words, around 18% of student-players will obtain 6 correct answers and 14 incorrect answers, and the same percentage will obtain 7 right and 13 wrong answers. With these, the most common results, and there being no doubt at this point that in a test of this kind the correct answer is worth $\frac{1}{2}$ points and the incorrect answer, $\frac{-1}{4}$ points, the following final marks will be obtained, respectively:

$$\text{Score with 6 hits} = 6 \cdot \frac{1}{2} \quad -14 \cdot \frac{1}{4} = -0{,}5$$
$$\text{Score with 7 hits} = 7 \cdot \frac{1}{2} \quad -13 \cdot \frac{1}{4} = 0{,}25$$

Conversely, being so unlucky as not to randomly answer any of the 20 questions correctly is a circumstance that only afflicts the

$$P(X = 0) \cdot 100 = \binom{20}{0} p^0 q^{20} \cdot 100 = q^{20} \cdot 100 = 0.0003007287 \cdot 100 \approx 0.03\%$$

of the student-players or, alternatively, of the times when a student takes a gamble, guessing the answers. At the other extreme, getting all the questions right by sheer luck is an event that has an even smaller associated probability, of only (expressed as a percentage),

$$P(X = 20) \cdot 100 = \binom{20}{20} p^{20} q^0 \cdot 100 = p^{20} \cdot 100 = 2.867972 \cdot 10^{-10} \cdot 100 = 2.867972 \cdot 10^{-8} = 0.0000000286792\%$$

in other words, it is an event that we expect to occur with a frequency of around three times every hundred million attempts.

Another relevant question concerns the minimum number of correct answers necessary to pass the test with the given characteristics ($n = 20, m = 3$), i.e., to obtain 5 or more points, which is the mark conventionally interpreted as a "pass" in exams with a maximum mark of 10. Thus:

$$x \cdot \frac{1}{2} \quad -(20 - x) \cdot \frac{1}{4} \geq 5$$
$$\frac{x}{2} \quad \frac{-20 - x}{4} \geq 5$$
$$\frac{2x - 20 + x}{4} \quad \geq 5$$
$$3x - 20 \quad \geq 20$$
$$3x \quad \geq 40$$
$$x \geq \frac{40}{3} \quad \geq 13{,}\overline{3}$$

Therefore, with a test of these characteristics, the student needs to get 14 answers right, since they will still score less than 5 with 13 correct answers:

$$\text{Score with 13 hits} = 13 \cdot \frac{1}{2} \quad -7 \cdot \frac{1}{4} = 4{,}75$$
$$\text{Score with 14 hits} = 14 \cdot \frac{1}{2} \quad -6 \cdot \frac{1}{4} = 5{,}5$$

The probability of obtaining the 14 correct answers necessary to pass the test is

$$P(X = 14) = \binom{20}{14} p^{14} q^{6} = 0.0007114406$$

In other words, 0.07%: seven out of ten thousand students or attempts will succeed.

In comparison, if there were no penalty for the wrong answer, answering half of the questions correctly would be enough to pass, with students safe in the comfort that they would be able to try their luck on all the questions. In other words, it would be possible to pass the exam even if half of the questions were answered incorrectly. The probability of this happening in an exam with 20 questions with three answer options each is

$$P(X = 10) = \binom{20}{10} p^{10} q^{10} = 0.0542592$$

Around five student-players out of a hundred would pass the test. Note that the order of magnitude of this number of passes is one hundred times higher than when incorrect answers are penalised and the student answers all the questions.

6. Conclusion

This methodological note includes a main conclusion for any teacher who uses a test to evaluate their students. Without making use of the unintuitive mathematical notation and its index or dummy variables (m, n, etc.), the conclusion can be expressed as follows: the correct answer must be worth...

$$\frac{\text{Maximum test score}}{\text{Number of test questions}} \text{ points}$$

The wrong answer must be worth...

$$\frac{\text{Maximum test score}}{\text{Number of test questions}} \cdot \frac{\text{-1}}{\text{Number of options per question} - 1} \text{ points}$$

Clearly explaining this scoring scheme to students, emphasising the low probability of passing the test by pure chance, will modify the "definition of the situation" (Thomas and Thomas, 1928) and this, in turn, will influence the student's preparation: the test is no longer a fertile social context for the appearance of rational players who make the most of their probability of answering correctly at random.

However, it is worth concluding this article by drawing attention to the fact that, as Pes (2009) explains, the strategy of the student-player can be extended to the calculation of the optimal number of questions to answer when wrong answers are penalised. Following his advice, there are students who access repositories with all the tests taken to date in their subject (past UNED tests, for example, are available to members of its student body; see UNED, 2020). Thus, they study by memorising the questions that have historically appeared most frequently in the different exam sittings. If, on the day of the exam, enough of these memorised questions appear to make the pass mark, the best strategy is to not answer any other questions.

To avoid this subterfuge used not by the model student-player, but rather by the student-strategist who seeks to pass with the minimum effort, memorising mechanically and without mastering the content of the course, the teacher could encourage students to answer all the questions by penalising both the unanswered question and the incorrect answer equally. After all, if the assumption is made that each question has a correct answer, not answering a question is itself a wrong answer. In summary, the teacher could counteract the student-strategist by scoring a display of knowledge positively and a lack of knowledge negatively, represented by both an unanswered question and the incorrect answer. This grading scheme also addresses the concern of Muñoz Clares and Caballero Salinas (2019) regarding the differential treatment of unanswered questions and wrong answers (legal basis of the campaign to prohibit penalisation in the tests promoted by Icaro100, 2010).

In short, my proposal would establish a new definition of the situation where the student is aware of the need to study the course content seriously. Doing so will increase their probability of knowing the correct answer up to the maximum $p = 1$ (absolute certainty), or up to an amount p higher than that which corresponds to pure chance. This is because those students who live up to their name and study will be able to restrict the possible correct answers to a subset of the original options. As Dehnad, Nasser and Hosseini (2014) recall, this situation, favoured by tests with three options per question where the conjecture is restricted to two options, no longer corresponds to an act of random guessing. It is now an "informed guess" with a final degree of success in direct relation to the intensity of study and the understanding of the subject. In other words, the multiple-choice exam is close to its ideal as an instrument to numerically assess the degree of knowledge achieved by the student.

A question related to the topic of this article that remains open for future contributions is posed by the different formats of multiple-choice questions that exist in addition to those with a single correct response option. Case and Swanson (2001) offer a systematic treatment of these alternatives, among which those questions with multiple correct response options stand out (Palés-Argullós, 2010). Following Gaviria's (2020) approach to the question, consideration might be given to how to score these other alternative test formats so that the student-player's expectation remains at zero.

7. Acknowledgements

I am grateful to the two anonymous reviewers from the *CENTRA Journal*, as well as my colleagues Verónica de Miguel and Ricardo Mora, for their comments on a preliminary draft of this article. Any errors and typos that, no doubt, unfortunately remain are my sole responsibility.

8. References

American Statistical Association (2013). *Should you guess? Interview with Rajeeva Laxman Karandikar*. Recovered on July 8, 2023 from http://www.worldofstatistics.org/2013/03/04/should-you-guess/.

Andreasen, K. E., Rasmussen, A. & Ydesen, C. (2013). Standardized testing. In J. Ainsworth (Ed.), *Sociology of Education: An A-to-Z Guide* (pp. 737-739). Thousand Oaks, CA: Sage.

Arnáiz, G. (1986). *Introducción a la estadística teórica*. Madrid: Lex Nova.

Baclawski, K. (2008). *Introduction to probability with R*. Boca Raton, FL: Chapman & Hall/CRC. https://doi.org/10.1201/9781420065220

Bennett de Marrais, K. & LeCompte, M. (1998). *The way schools work: A sociological analysis of education*. London: Longman/Addison Wesley.

Bernoulli, J. ([1713] 1993). *Teoría de probabilidades (Ars conjectandi).* Madrid: Sociedad Española de Historia de las Ciencias y de las Técnicas, Llull.

Bickis, Mikelis G. (2017). *STATS 244.3. Elementary Statistical Concepts. Scoring Multiple Choice Exams.* Acceso 8 julio 2023. From http://math.usask.ca/~bickis/courses/stats244/marking.html.

Case, S. M. & Swanson, D. B. (2001). *Constructing Written Test Questions for the Basic and Clinical Sciences*. 3rd edition. Philadelphia, PA: National Board of Medical Examiners.

Dehnad, A., Nasser, H. & Hosseini, A. F. (2014). A Comparison Between Three-and Four-Option Multiple Choice Questions. *Procedia. Social and Behavioral Sciences, 98*, 398-403. https://doi.org/10.1016/j.sbspro.2014.03.432

Edwards, N. T. (2006). The historical and social foundations of standardized testing: In search of a balance between learning and evaluation. *Shiken and Shiken Research Journal: JALT Testing and Evaluations SIG Newsletter, 10*, 8-16.

Gaviria, J. L. (2020). *Recomendaciones para la Elaboración de las Preguntas*. Vicerrectorado de Tecnología y Sostenibilidad, Universidad Complutense de Madrid. Acceso 8 julio 2023. Disponible en: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiBoYOztv__AhWEXcAKHf-nAIcQFnoECA4QAQ&url=https%3A%2F%2Fderecho.ucm.es%2Ffile%2Frecomendaciones-para-elaborar-preguntas-a-traves-de-herramienta-cuestionario-moodle&usg=AOvVaw0cZadEz-i2iJ-rQwHGdZsp&opi=89978449.

Goslin, D. A. & Glass, D. C. (1967). The social effects of standardized testing in American elementary and secondary Schools. *Sociology of Education, 40* (2), 115-131. https://doi.org/10.2307/2112040

Icaro100 (2010). Campaña de impugnación de exámenes tipo test alegando la ilegalidad de la penalización. *Tribuna del Jurista.* Recovered on July 8, 2023 from: https://tribunadeljurista.foroes.org/t1486-campaa-de-impugnacin-de-exmenes-tipo-test-alegando-la-ilegalidad-de-la-penalizacin-adjunto-formulario-de-impugnacin.

Levitt, S. D. & List, J. A. (2008). Homo economicus evolves. *Science, 319* (5865), 909-910. https://doi.org/10.1126/science.1153640

Martín-Pliego, F. J. & Ruiz-Maya, L. (2006). *Fundamentos de Probabilidad.* Madrid: Ediciones Paraninfo.

Morales, M. A. (2017). Las matemáticas de la fórmula de puntuación de exámenes test. *El País,* 22 de febrero. Recovered on July 8, 2023 from: https://elpais.com/elpais/2017/02/22/el_aleph/1487760836225306.html.

Muñoz Clares, J. & Caballero Salinas, J. M. (2019). Problemática de los exámenes test con merma por respuesta errónea. *Revista Aranzadi Doctrinal*, 7.

Palés-Argullós, J. (2010). ¿Cómo elaborar correctamente preguntas de elección múltiple? *Educación Médica*, 3. https://doi.org/10.4321/S1575-18132010000300005

Pes, C. (2009). *¿Cómo Aprobar Un Examen Tipo Test?* Recovered on July 8, 2023 from: https://www.carlospes.com/articulos/como_aprobar_un_examen_tipo_test.php.

Psiconociendo (2022). *12 Trucos Para Aprobar Un Examen Tipo Test Sin Estudiar*. Youtube. Recovered on July 8, 2023 from: https://www.youtube.com/watch?v=gR9yHCOW05c&list=RDLVuv0hBxbx3Tk&index=6.

Schneid, S. D., Armour, C., Park, Y. S., Yudkowsky, R. & Bordage, G. (2014). Reducing the number of options on multiple-choice questions: Response time, psychometrics and standard Setting. *Medical Education, 48* (10): 1020-1027. https://doi.org/10.1111/medu.12525

Sentipia (2022). *Cómo Responder Exámenes Tipo Test para Aprobar.* Youtube. Recovered on July 8, 2023 from: https://www.youtube.com/watch?v=uv0hBxbx3Tk&list=RDLVuv0hBxbx3Tk&index=1.

Stanbrough, J. L. (2009). *The guessing penalty. Let's do the math, shall we?* Batesville, IN: Batesville Community School Corporation. Recovered on July 8, 2023 from: http://www.batesville.k12.in.us/Physics/CalcNet/Intro/guessing_penalty.htm

Thomas, W. I. y Thomas, D. S. (1928). *The Child in America: Behavior Problems and Programs*. New York, NY: Knopf.

UNED (2020). *Acceso al Depósito de Exámenes*. Recovered on July 8, 2023 from: http://www.calatayud.uned.es/examenes/examenes_step_0.asp.

Weber, Max ([1922] 1982). *Ensayos sobre metodología sociológica.* Buenos Aires: Amorrortu.

Notes

1 With this case study presented in this article, the empirical result offered by the mean of a data set approximates the value of the expectation of the probability distribution when (a) the same student-player randomly takes a gamble in the exam n times, with $n \to \infty$; or (b) there are a number k of students who answer the exam question at random, with $k \to \infty$.

2 The Bernoulli distribution is the particular case of the binomial distribution for $n = 1$ tests. Subsequently, $X \sim \text{Bernoulli}\ (p) = B\ (1, p)$ and, therefore, the random variable is stated as success/correct answer (vs. failure/incorrect answer) and, implicitly, it is meant to mean "in a single trial or experiment".

The author

Daniel Guinea-Martín is a lecturer in Sociology at the University of Malaga where he teaches methods, sociology of education and tourism. He holds a PhD from the European University Institute of Florence, a master's degree from the London School of Economics, a diploma from the University of California, San Diego, as well as a degree in sociology and a degree in statistics from the Complutense University of Madrid. He specialises in social research methods, inequality and segregation. His work has been published in journals such as the American Sociological Review, Social Science Research and Ethnography, among others.