Journal of Uncertain Systems
Vol.8, No.3, pp.216-221, 2014
Online at: www.jus.org.uk

"Fuzzy" Multiple-Choice Quizzes and How to Grade Them

Shahnaz Shahbazova[1,*], Olga Kosheleva[2]
[1] *Azerbaijan Technical University, Baku, Azerbaijan*
[2] *Department of Teacher Education, University of Texas at El Paso, El Paso, TX 79968, USA*

Received 22 August 2013; Revised 24 August 2013

Abstract

Traditional multiple-choice quizzes do not tell us how confident the students are in their answers. To provide this additional information, researchers have proposed "fuzzy" multiple-choice quizzes, in which a student explicitly describes his/her degree of confidence in each possible answer. The idea is reasonable, but, as we show, the current grading scheme for such quizzes discourages students from providing the correct degrees of confidence. In this paper, we show that there is only one grading scheme which encourages students to provide the correct degrees: when the number of points is proportional to the logarithm of the degree. Somewhat surprisingly, this scheme turns out to be related to the notion of entropy.

Keywords: fuzzy multiple-choice quizzes, grading, entropy

1 What are "Fuzzy" Multiple-Choice Quizzes

Multi-choice quizzes are actively used. A typical multiple-choice quiz consists of several questions. For each question, several possible answers are presented to the student; a student selects one of these answers. The number of correctly answered questions determines the student's grade.

These quizzes are frequently used, since:

- they do not take too much time to administer, and
- they are easy to grade – in particular, they are easy to grade automatically.

Limitations of the multiple-choice quizzes. Multiple-choice quizzes are not a panacea, they have known limitations:

- first, they can be applied only when the answer is clear-cut, e.g., when we are checking a historical fact or an answer to a well-defined arithmetic problem;
- second, if the answer, e.g., to an arithmetic problem is wrong, we do not know whether it is because the student has not yet learned how to solve such problems or because he or she made a simple arithmetic mistake (in which case, in a normal exam, this student would get a partial credit).

To overcomes these two limitations, we have to administer a "real" exam, in which students are required to provide detailed answers to their questions.

There is another limitation to multiple-choice quizzes: we do not know how confident are the students about their answers. For example, if 72% of the students provided a correct answer to a 5-choice question, this fact can have two different explanations:

- It can be that 72% of the students confidently know the correct answer. With an often-used 70% threshold for satisfactory learning, this would mean that the students have satisfactorily mastered this particular question.

*Corresponding author.
Emails: shahbazova@gmail.com (S. Shahbazova), olgak@utep.edu (O. Kosheleva).

- Alternatively, this could also mean that only 65% of the students know the correct answer, while the other 35% have chosen one of the five answers at random. In this case, the total number of correct answer is also equal to 65% + (35%/5) = 72%; however, the number of students who know the correct answer (65%) is below the satisfactory level of 70%.

Main idea behind "fuzzy" multiple-choice quizzes. As a part of the team-based learning paradigm (see, e.g., [5, 6, 7, 11]), researchers have proposed a modification of the usual multiple-choice quizzes which allows the students to also describe their degree of confidence in different possible answers.

Specifically, for each question, instead of asking a student to mark one of the options (e.g., a, b, c, or d), the researchers propose to have *several* slots so that:

- a student who is confident that a is the correct answer can mark $a\ a\ a\ a$ in all these slots, while
- a student who is not so sure can mark $a\ a\ a\ b$.

This way, the second student can get some partial credit if b is the correct answer – although, of course, he or she get fewer points if the correct answer is a.

Relation to fuzzy. In effect, the above technique means that instead of selecting a crisp answer i ($1 \le i \le n$), a student can have a "fuzzy" answer, with different weights $w_i \ge 0$ given to different answers i, weights for which $\sum\limits_{i=1}^{n} w_i = 1$.

This idea is in the spirit of *fuzzy logic*; see, e.g., [2, 8, 12]. Specifically, is similar to *fuzzy voting* idea (described, e.g., in [3]) where instead of voting for a single candidate a voter can assign weights $w_i \ge 0$ $\left(\sum\limits_{i=1}^{n} w_i = 1\right)$ to different candidates.

How "fuzzy" multiple-choice quizzes are graded now. At present, the grade (number of points) given to a student is proportional to the weight w_i that this student assigns to the correct answer i: $g = g_0 \cdot w_i$ for some constant g_0.

- When the student is absolutely sure about the correct answer and assigns weight $w_i = 1$ to this answer (and $w_j = 0$ to all other answers), this student gets the maximum number of points c.
- on the other hand, if the student only assigns the weight $w_i = 0.5$ to the correct answer, this student gets half of the possible number of points, i.e., $0.5 \cdot g_0$.

2 Problem with How "Fuzzy" Multiple-Choice Quizzes are Graded Now

Discussion. At first glance, the above grading scheme sounds very reasonable. However, as we will show, this grading scheme does not always encourage a student to provide the correct weights w_i. To show this, let us describe, in precise terms, how a student will select the weights.

How to describe the student's uncertainty. According to decision theory, we can describe a student's uncertainty by assigning (subjective) probabilities p_i to different answers, so that $\sum\limits_{i=1}^{n} p_i = 1$; see, e.g., [1, 4, 9].

Ideal selection of weights. Ideally, we would like a student to select $w_i = p_i$.

However, as we will see, this is not what will happen for the above grading scheme.

Objective function. A student is answering many questions on many quizzes. The student's objective is to gain as many points on all these quizzes as possible. In other words, a student needs to maximize the sum of the grades gained on all the questions of all the quizzes.

Maximizing the *sum* is equivalent to maximizing the *average*. Thus, on each question, the student selects the weights so as to maximize the average number of points.

In general, the number of points depends on the weight w_i that a student gave to the correct answer i. Let us denote the corresponding number of points by $g(w_i)$. In the above grading scheme, $g(w_i) = c \cdot w_i$.

Based on the student's knowledge, the probability that i is the correct answer is p_i. In i is the correct answer, the student gets $g(w_i)$ points. Thus, the total average (expected number) of points is equal to $\sum\limits_{i=1}^{n} p_i \cdot g(w_i)$.

Resulting optimization problem. Thus, when a student selects the weights, he or she solves the following optimization problem:

- we are *given* the probabilities $p_1, \ldots, p_n \geq 0$ for which $\sum\limits_{i=1}^{n} p_i = 1$;
- we are also *given* a function $g(w)$;
- we *need to find* the weights $w_1, \ldots, w_n$ which maximize the sum $\sum\limits_{i=1}^{n} p_i \cdot g(w_i)$ under the constraint $\sum\limits_{i=1}^{n} w_i = 1$.

Problem with the usual grading scheme. In the usual grading scheme, $g(w) = g_0 \cdot w$. For this scheme, the objective function takes the form $\sum\limits_{i=1}^{n} g_0 \cdot p_i \cdot w_i$. Let us show that for this objective function, its maximum is attained when we assign the full weight $w_{i_0} = 1$ to the alternative i_0 with the largest probability $p_{i_0} = \max\limits_i p_i$ (and we assign 0 weights $w_i = 0$ to all other possible answers $i \neq i_0$).

Indeed, for this selection, the resulting value of the objective function is $c \cdot p_{i_0} = g_0 \cdot \max\limits_i p_i$. On the other hand, since $p_i \leq p_{i_0}$ for all i, for every other selection of weights, we have

$$\sum_{i=1}^{n} g_0 \cdot p_i \cdot w_i \leq g_0 \cdot p_{i_0} \cdot \sum_{i=1}^{n} w_i = g_0 \cdot p_{i_0}.$$

Thus, a student will always prefer a crisp assignment (except for the case when two or more possible answers have the same probability).

So, for the above grading scheme, a rational student will select one of the alternatives with weight 1 and will, therefore, *not* reveal his or her degree of uncertainty – which defeats the whole purpose of the "fuzzy" multiple-choice quiz.

What we do. In this paper, we propose an alternative grading scheme which encourages students to reveal their degree of uncertainty. Moreover, we prove that there is only one such grading scheme.

3 Towards a Grading Scheme that Encourages Students to Reveal Their Degrees of Certainty

Formulation of the problem. We want to find a grading scheme, i.e., a function $g(w)$, for which the students should always select the weights $w_i = p_i$.

Formulation of the problem in precise terms. To find the desired grading scheme, let us reformulate this problem in precise terms.

We want to find a function $g(w)$ for which, for each tuple $p_1, \ldots, p_n \geq 0$ with $\sum\limits_{i=1}^n p_i = 1$, the maximum of the expression $\sum\limits_{i=1}^n p_i \cdot g(w_i)$ under the constraint $\sum\limits_{i=1}^n w_i = 1$ is attained for $w_i = p_i$.

Analysis of the problem. The above formulation means that if we deviate from $w_i = p_i$ while preserving the constraint $\sum\limits_{i=1}^n w_i = 1$, the value of the objective function $\sum\limits_{i=1}^n p_i \cdot g(w_i)$ can only decrease.

Let us start with the ideal assignment $w_i = p_i$, and let us consider what happens when we keep $w_i = p_i$ for all $i > 2$ and change the first two original values $w_1 = p_1$ and $w_2 = p_2$ to new values w_1 and $w_2 = p - w_1$, where we denoted $p \stackrel{\text{def}}{=} p_1 + p_2$. After this change, the objective function takes the form

$$p_1 \cdot g(w_1) + p_2 \cdot g(p - w_1) + \sum_{i=3}^n p_i \cdot g(p_i).$$

This expression should attain its largest possible value when $w_1 = p_1$; thus, for $w_1 = p_1$, the derivative of this expression with respect to w_1 should be equal to 0. Differentiating the above expression, substituting $w_1 = p_1$, and equating the derivative to 0, we conclude that

$$p_1 \cdot g'(p_1) - p_2 \cdot g'(p - p_1) = 0,$$

i.e., since $p - p_1 = (p_1 + p_2) - p_1 = p_2$, that

$$p_1 \cdot g'(p_1) = p_2 \cdot g'(p_2).$$

This equality must be true for all possible values of p_i, so we conclude that

$$p \cdot g'(p) = C$$

for some constant C. From $g'(p) = C/p$, by integrating both sides, we conclude that $g(p) = C \cdot \ln(p) + C_1$, where C_1 is an integration constant.

Comment. Of course, we want to assign more points to more confident students, so we should take $C > 0$.

Conclusion. For a grading scheme to encourage students to reveal their true degrees of confidence, we must select a grading function $g(w)$ which is proportional to the logarithm of the weight.

In other words, the number of points assigned to each student's answer must be proportional *not* to the weight w_i which a student have to the correct answer but to the *logarithm* $\ln(w_i)$ of this weight.

Double-checking our solution. Strictly speaking, we have only shown that *if* the grading scheme leads to the ideal weight assignment $w_i = p_i$, then we should have $g(w)$ proportional to $\ln(w)$. Let us show that, vice versa, if we select a logarithmic grading function $g(w)$, then we should always get $w_i = p_i$.

Indeed, by computing the second derivatives with respect to w_i, one can see that for the objective function $f(w_1, \ldots, w_n) \stackrel{\text{def}}{=} \sum\limits_{i=1}^n p_i \cdot (C \cdot \ln(w_i) + C_1)$, we have

$$\frac{\partial^2 f}{\partial w_i^2} = -\frac{C \cdot p_i}{w_i} < 0$$

and

$$\frac{\partial^2 f}{\partial w_i \partial w_j} = 0$$

for $i \neq j$. Thus, the objective function is concave – and its restriction to the plane $\sum\limits_{i=1}^n w_i = 1$ is concave too; therefore, its maximum is attained at a single point $(w_1, \ldots, w_n)$; see, e.g., [10].

This constraint maximum can be found by using Lagrange multiplier method. According to this method, the maximum of the original constraint optimization problem coincides, for the appropriate value of the Lagrange multiplier λ, with the unconstrained maximum of the auxiliary function

$$F(w_1, \ldots, w_n) \stackrel{\text{def}}{=} \sum_{i=1}^{n} p_i \cdot (C \cdot \ln(w_i) + C_1) + \lambda \cdot \left(\sum_{i=1}^{n} w_i - 1 \right).$$

Differentiating this auxiliary function by w_i and equating the derivative to 0, we conclude that $C \cdot p_i/w_i + \lambda = 0$, i.e., that $p_i/w_i = c \stackrel{\text{def}}{=} -\lambda/C$, and $w_i = c^{-1} \cdot p_i$. Since $\sum\limits_{i=1}^{n} p_i = 1$ and $\sum\limits_{i=1}^{n} w_i = 1$, we conclude that $c^{-1} = 1$ and thus, $w_i = p_i$. The statement is proven.

How to select the parameters C and C_1. We would like to select the constants C and C_1 in the expression $g(w) = C \cdot \ln(w) + C_1$ in such a way that the largest number of points for this quiz is equal to g_0 and the smallest number of points is equal to 0.

A seemingly natural idea is to find C and C_1 for which $g(0) = 0$ and $g(1) = g_0$; however, this is not possible, since $g(0) = -\infty$. This is not a serious problem, however, since $w = 0$ means that a student is absolutely confident in the wrong answer; this is rarely the case.

Let us therefore set the lower bound w_0 on the weights, so that weights $w_i < w_0$ will still be counted as w_0. In this case, we need to select C and C_1 in such a way that $g(w_0) = 0$ and $g(1) = g_0$, i.e., that $C \cdot \ln(w_0) + C_1 = 0$ and $C \cdot \ln(1) + C_1 = g_0$. The second equality leads to $C_1 = g_0$, and the first one leads to $C = -g_0/\ln(w_0)$. Thus,

$$g(w) = g_0 \cdot \left(1 - \frac{\ln(w)}{\ln(w_0)}\right) = g_0 \cdot \frac{\ln(w_0) - \ln(w)}{\ln(w_0)} = \frac{g_0}{|\ln(w_0)|} \cdot \ln\left(\frac{w}{w_0}\right).$$

Final formula. When a student assigns grade w_i to the correct answer, this student should get the following number of points:

$$g(w_i) = \frac{g_0}{|\ln(w_0)|} \cdot \ln\left(\frac{w_i}{w_0}\right).$$

Numerical example. When a student has four slots to fill, the smallest probability that can be thus taken into consideration is $1/4$, so it makes sense to take $w_0 = 1/8$. In this case, $\ln(w_0) = -\ln(8)$, so $|\ln(w_0)| = \ln(8)$.

- If a student marks all four slots with the correct answer, i.e., if $w_i = 1$, then this student gets the full grade.

- If a student marks three out of four slots with the correct answer, i.e., if $w_i = 3/4$, then the student gets the portion of
$$\frac{\ln(6)}{\ln(8)} \approx 0.86$$
of the full grade.

- If a student marks two out of four slots with the correct answer, i.e., if $w_i = 1/2$, then the student gets the portion of
$$\frac{\ln(4)}{\ln(8)} = \frac{2}{3} \approx 0.67$$
of the full grade.

- If a student marks one out of four slots with the correct answer, i.e., if $w_i = 1/4$, then the student gets the portion of
$$\frac{\ln(2)}{\ln(8)} = \frac{2}{3} \approx 0.33$$
of the full grade.

- Finally, if a student does not mark any of four slots with the correct answer, i.e., if $w_i = 0$, then we take $w_i = w_0$, and thus, the student gets
$$\frac{\ln(1)}{\ln(8)} = 0$$
points.

What is the resulting expected grade? An interesting relation to entropy. When a student has uncertainty described by the probabilities $p_1, \ldots, p_n$, and marks $w_i = p_i$, what is the expected (average) grade f that this student will get?

Substituting $w_i = p_i$ and the expression $g(w_i) = C \cdot \ln(w_i) + C_1$ into the objective function $f = \sum\limits_{i=1}^{n} p_i \cdot g(w_i)$, we conclude that $f = C \cdot \sum\limits_{i=1}^{n} p_i \cdot \ln(p_i) + C_1$. Interestingly, this expression is related to the *entropy* $S \stackrel{\text{def}}{=} -\sum\limits_{i=1}^{n} p_i \cdot \ln(p_i)$ of the corresponding probability distribution, as

$$f = C_1 - C \cdot S = g_0 \cdot \left(1 - \frac{S}{|\ln(w_0)|}\right).$$

Acknowledgments

The authors are thankful to Dr. Laura Madson for presenting an inspiring workshop on team-based learning, and to Vladik Kreinovich and Larry Lesser for stimulating discussions.

References

[1] Fishburn, P.C., *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.

[2] Klir, G.J., and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[3] Kosko, B., *Heaven in a Chip: Fuzzy Visions of Society and Science in the Digital Age*, Three Rivers Press, New York, 2000.

[4] Luce, R.D., and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[5] Michaelsen, L.K., Knight, A.B., and L.D. Fink (eds.), *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*, Stylus Publishing, Sterling, Virginia, 2004.

[6] Michaelsen, L.K., Parmelee, D.X., McMahon, K.K., and R.E. Levine (eds.), *Team-Based Learning for Health Professions Education: A Guide to Using Small Groups for Improving Learning*, Stylus Publishing, Sterling, Virginia, 2007.

[7] Michaelsen, L.K., Sweet, M., and D.X. Parmelee (eds.), *Team-Based Learning: Small Group Learning's Next Big Step. New Directions for Teaching and Learning*, Jossey-Bass, San Francisco, California, 2008.

[8] Nguyen, H.T., and E.A. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[9] Raiffa, H., *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[10] Rockafeller, R.T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[11] Sweet, M., and L.K. Michaelsen (eds.), *Team-Based Learning in the Social Sciences and Humanities: Group Work that Works to Generate Critical Thinking and Engagement*, Stylus Publishing, Sterling, Virginia, 2012.

[12] Zadeh, L.A., Fuzzy sets, *Information and Control*, vol.8, pp.338–353, 1965.