How to Make Sure that the Grading Scheme Encourages Students to Learn All the Material: Fuzzy-Motivated Solution and Its Justification

Olga Kosheleva

Abstract— **In many practical situations, it is desirable that the students learn all the parts of the material. It is therefore desirable to set up a grading scheme that encourages such learning. We show that the usual scheme of computing the overall grade for the class – as a weighted average of grades for different assignments and exams – does not always encourage such learning. Each such intermediate grade describes the student's knowledge of a certain part of the material. From the viewpoint of fuzzy logic, the degree to which the student knows the 1st part of the material *and* the 2nd part of the material, etc., can be naturally described as a result of applying a t-norm ("and"-operation) to intermediate degrees (intermediate grades) – e.g., as the minimum of the intermediate grades. It turns out that this fuzzy-motivated** $\min$ **grading scheme indeed encourages students to learn all the material – and vice versa, the only grading scheme that provides such encouragement is the minimum of the intermediate grades.**

I. Formulation of the Problem

It is often important that the students learn all the material. The material taught in a typical semester-long class consists of several parts. In many cases, it is important that a student get reasonable knowledge of all the parts of the material. This is clear for such disciplines as medicine – we want a medical doctor to have basic knowledge of all types of diseases – but is also important in many other disciplines as well.

The desired level of knowledge may be different in different applications. For example, a medical doctor who just starts his internship under the mentorship of a skilled professional may have satisfactory knowledge of some parts, since the mentor is there to help. On the other hand, when the doctor is certified as capable to start his or her own medical practice, we would like the doctor to have good knowledge of *all* parts of the required material.

The grading scheme should reflect this requirement. It is desirable that the grading scheme not only gauge how well the students learn the material; the grading scheme should also encourage the students to learn *all* the parts of the material.

Towards formalizing this idea: how a student plans his or her studies. A student has a limited time t that can allocated to learning the material. This time needs to be distributed between n different parts of the material, i.e., the student must select, for each part $i = 1, 2, \ldots, n$, the time $t_i \geq 0$ that is allocated for studying this part. The selected times t_i should add up to the given amount t:

$$t_1 + t_2 + \ldots + t_n = t.$$

How to quantify knowledge. For each part of the material, it is reasonable to describe the student's knowledge in terms of a proportion of the material that the student learned, i.e., by a number from 0 to 1 such that 0 means no knowledge at all, and 1 means perfect knowledge. This can be estimated, e.g., as the proportion of correct answers on a comprehensive exam.

How student learn. Let us assume that for each value $t \geq 0$, we know the amount of knowledge $a(t)$ that a student will achieve if he or she studies the corresponding part for time t.

The more time the student learns, the more knowledge he or she acquires – unless the student already achieved the perfect knowledge $a(t) = 1$. In mathematical terms, this means that the function $a(t)$ is (non-strictly) increasing:

$$\text{if } t \leq t', \text{ then } a(t) \leq a(t').$$

It is also reasonable to assume that if a student slightly changes the time amount, the resulting knowledge will also change only slightly. In mathematical terms, this means that the function $a(t)$ is continuous.

The function $a(t)$ may differ from one group of students to others: some students have steeper learning curve, some learn slowly, etc.

Grading: general idea. Let us also assume that the tests, labs, and home assignments correctly gauge this knowledge. As a result, for each part of the material, we know the level of knowledge $a_i \stackrel{\text{def}}{=} a(t_i)$ acquired by the student i.

A grading scheme is a method F that transforms the n values $a_1, \ldots, a_n$ describing the student's knowledge of different parts of the material into a single grade

$$a = F(a_1, \ldots, a_n).$$

If all the grades a_i are the same, i.e., if

$$a_1 = a_2 = \ldots = a_n,$$

then it is reasonable to take this common grade as the grade for the class, i.e., to assume that

$$F(a, \ldots, a) = a.$$

Olga Kosheleva is with the Department of Teacher Education, University of Texas at El Paso, 500 W. University, El Paso, TX9968, USA (email: olgak@utep.edu).

In mathematical terms, this means that the function F should be *idempotent*.

The more students know about each part of the material, the better should be the overall grade. In mathematical terms, this means that the function F should be monotonic, i.e.:

if $a_i \le a'_i$ for all i, then we should have

$$F(a_1, \ldots, a_n) \le F(a'_1, \ldots, a'_n).$$

It is also reasonable to require that if a_i changes a little bit, the resulting grade should not change much. In precise terms, this means that the function F should be continuous.

How grading is done now. Usually, the overall grade is computed as the weighted average of grades corresponding to different parts of the material:

$$F(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i \cdot a_i$$

for some weights $w_i \ge 0$ for which

$$\sum_{i=1}^{n} w_i = 1.$$

The weighted average function is clearly monotonic and continuous.

How students decide how much time to allocate to different topics. When a student allocates time t_i to topic i, the student's grade for topic i becomes $a_i = a(t_i)$, and his or her overall grade is equal to

$$a = F(a(t_1), \ldots, a(t_n)).$$

The student wants to maximize this grade, i.e., select the allocations $t_1, \ldots, t_n$ for which this value

$$F(a(t_1), \ldots, a(t_n))$$

is the largest possible.

What we want. Let us denote by a_0 the desired degree of knowledge in every topic. We then want to make sure that the grading scheme (i.e., the function F) is such that if it is possible to find time allocation for which $a(t_i) \ge t_0$ for all i, then the allocation selected by the student will satisfy this property.

Ideally, this should be true for all types of students, with different functions $a(t)$.

II. The Desired Quantity Is Not Always Satisfied for the Current Grading System

Indeed, suppose that we have several parts of the material, and we want to get a level a_0 on all these parts – e.g., a satisfactory level $a_0 = 0.7$. Suppose also that the student's learning curve is $a(t) = t^2$ (describing a steep learning curve). In this case, to achieve the desired level of knowledge on each of n topics, the student needs to spend time t_0 for which $t_0^2 = a_0$, i.e., time $t_0 = \sqrt{a_0}$. After spending this amount of time on each of n topics, the student spends the total time $t = n \cdot t_0 = n \cdot \sqrt{a_0}$.

In this case, the student earns the grade a_0 on all the topics. Let us assume that all n grades are equally weighted, i.e., the overall grade is the arithmetic average of all n grades $a_i = a(t_i)$. In this case, for the desired student behavior, we get the overall grade

$$g = \frac{a_0 + \ldots + a_0}{n} = a_0.$$

Alternative, the student can spend this time by spending time $t_i = 1$ on $n \cdot \sqrt{a_0}$ out of n topics, and no time on remaining topics. For selected topics, this student get

$$a_i = a(t_i) = 1,$$

the perfect knowledge. For other topics, the student gets 0 knowledge

$$a_i = a(t_i) = a(0) = 0.$$

The overall grade is thus equal to

$$\frac{n \cdot \sqrt{a_0}}{n} = \sqrt{a_0}.$$

For $a_0 < 1$, we have $\sqrt{a_0} > a_0$, so the student prefers the alternative learning strategy.

III. Heuristic Idea Motivated by Fuzzy Logic

We want the student to know:

- the 1st part of the material *and*
- the second part *and*
- …*and*
- the n-th part.

For each i, we know the degree a_i to which the student knows the i-th part of the material. Thus, according to the fuzzy logic methodology (see, e.g., [1], [2], [3]), the degree a to which the desired requirement is satisfied can be found by applying a fuzzy "and"-operation (t-norm) to these degrees.

The requirement that $F(a, a) = a$ is satisfied only by one fuzzy "and"-operation – the minimum $\min(a, b)$ [1], [2]. If we use this "and"-operation, we get the grading scheme

$$a = \min(a_1, \ldots, a_n).$$

In the above example, the new grading scheme leads to the desired student behavior. Indeed, when the student spends the same amount of time $\sqrt{a_0}$ on each topic and get grades $a_1 = \ldots = a_n = a_0$, his overall grade – according to the new grading scheme – is $\min(a_0, \ldots, a_0) = a_0$.

Alternatively, if the students gets a perfect grade $a_i = 1$ on $n \cdot \sqrt{a_0}$ topics and $a_i = 0$ on all other topics, his or her overall grade is

$$\min(1, \ldots, 1, 0, \ldots, 0) = 0.$$

Since $0 < a_0$, the student will clearly prefer the desired learning strategy.

IV. Formal Definitions and the Main Result

What we do in this paper. In this paper, we show the above-described behavior of the min grading scheme is not accidental. Specifically, we prove two results:

- that if we use the fuzzy-motivated min grading scheme, then the student would always prefers to equally distribute effort between different topics – exactly what we want to achieve;
- second, we prove that min grading is the only grading scheme with this property.

To describe these results in precise terms, let us first define the problem in precise terms.

Definition 1. *We say that a function $a(t_1, \ldots, t_n)$ is* (non-strictly) increasing *if whenever $t_1 \le t'_1$, …, and $t_n \le t'_n$, we have*

$$a(t_1, \ldots, t_n) \le a(t'_1, \ldots, t'_n).$$

Definition 2. *By a* learning curve*, we mean a continuous non-strictly increasing function $a(t)$ from non-negative real numbers to the interval $[0, 1]$.*

Definition 3. *We say that a function $F(a_1, \ldots, a_n)$ is* idempotent *if for every a, we have $F(a, \ldots, a) = a$.*

Definition 4. *For every integer $n \ge 2$, by a n-*grading scheme*, we mean a continuous non-strictly increasing idempotent function $F : [0, 1]^n \to [0, 1]$.*

Comment. As an example of the n-grading scheme, we have considered the min grading scheme

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n).$$

Definition 5. *Let $t > 0$ be a positive real number and let $n \ge 2$ be an integer. By a (t, n)-*learning strategy*, we mean a tuple of non-negative values $t_1, \ldots, t_n$ for which*

$$t_1 + \ldots + t_n = t.$$

Comment. Not all (t, n)-learning strategies are always possible. For example, we may make sure that the students study in class. In this case, instead of the set of all possible (t, n)-learning strategies, we may be restricting ourselves to a *set* $\mathcal{S}$ of such strategies.

Definition 6. *Let $a(t)$ be a learning curve, let $n \ge 2$ be an integer, let $F(a_1, \ldots, a_n)$ be an n-grading scheme, let $t > 0$ be a positive real number, let $(t_1, \ldots, t_n)$ be a (t, n)-learning strategy, and let $a_0 > 0$ be a positive real number.*

- *For every i from 1 to n, by the* grade for the i-th assignment*, we mean the value $a(t_i)$.*
- *We say that the learning strategy is* uniformly a_0-successful *if $a(t_i) \ge a_0$ for all i.*
- *By an* overall grade*, we mean the value $F(a(t_1), \ldots, a(t_n))$.*
- *Let $\mathcal{S}$ be a set of (t, n)-leaning strategies. We say that this learning strategy from this set is* $(\mathcal{S}, F)$-optimal *if its overall grade is large than or equal to the overall grade of all other (t, n)-learning strategies from the set $\mathcal{S}$.*

Definition 7. *Let $n \ge 2$ be an integer, and let $F(a_1, \ldots, a_n)$ be an n-grading scheme. We say that this grading scheme* encourages students to learn all the material *if for every learning curve $a(t)$, for every two positive real numbers t and a_0, and for every set $\mathcal{S}$ of (t, n)-learning strategies,*

- *if, in the set $\mathcal{S}$, there exists a uniformly a_0-successful (t, n)-learning strategy,*
- *then every $(\mathcal{S}, F)$-optimal (t, n)-learning strategy is uniformly a_0-successful.*

Theorem. *For every integer $n \ge 2$:*

- *the* min *grading scheme*

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n)$$

encourages students to learn all the material;

- *vice versa, if an n-grading scheme $F(a_1, \ldots, a_n)$ encourages students to learn all the material, then it coincides with the* min *grading scheme.*

V. Proof of the Theorem

1°. Let us first prove that the min grading scheme encourages students to learn all the material, i.e., that if there exists a uniformly a_0-successful (t, n)-learning strategy, then every min-optimal learning strategy is uniformly a_0-successful.

Indeed, for a uniformly a_0-successful strategy, by definition, we have $a_i = a(t_i) \ge a_0$ for all i. Thus, the overall grade

$$a = F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n)$$

corresponding to this strategy also has the property $a \ge a_0$.

For the optimal strategy, the grade, by definition, is $\ge a$ and thus, $\ge a_0$. So, for this strategy $(t_1, \ldots, t_n)$, we have $\min(a(t_1), \ldots, a(t_n)) \ge a_0$. Since each of n numbers $a(t_i)$ is larger than or equal to the smallest of them $\min(a(t_1), \ldots, a(t_n))$, we thus conclude that $a(t_i) \ge a_0$ for all i – i.e., the optimal learning strategy is indeed uniformly a_0-successful.

2°. Let us now prove the second part of our theorem. Let us assume that a grading scheme $F(a_1, \ldots, a_n)$ encourages students to learn all the material. Let us prove that in this case, the grading scheme F coincides with the min grading scheme, i.e., that

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n).$$

2.1°. It is sufficient to prove the above formula for the case when all the values a_i are positive.

Indeed, once we prove this formula for all positive a_i, we can use continuity to extend it to the case when some of the values a_i are equal to 0.

In view of this observation, in the remaining part of Part 2 of this proof, we will assume that $a_i > 0$ for all i.

2.2°. Let us prove that for every $m > 0$, for every $\varepsilon \in (0, m)$, and for every integer i from 1 to n, we have

$$F(1, \ldots, 1 \ (i-1 \text{ times}), m-\varepsilon, 1 \ldots, 1) < m.$$

Indeed, let us consider the function $a(t)$ for which:

- $a(0) = 0$,
- $a(1-\varepsilon) = m - \varepsilon$,
- $a(1) = m$, and
- $a\left(1 + \dfrac{\varepsilon}{n-1}\right) = 1$.

Let us also assume that that the function $a(t)$ is piece-wise linear, i.e., that it is linear on the intervals $(0, 1-\varepsilon)$, $(1-\varepsilon, 1)$, and $\left(1, 1 + \dfrac{\varepsilon}{n-1}\right)$. For $t \ge 1 + \dfrac{\varepsilon}{n-1}$, we have $a(t) = 1$. One can easily check that this function $a(t)$ is continuous and non-strictly increasing.

Let us assume that the threshold is $a_0 = m$. If we take

$$t_1 = \ldots = t_n = 1,$$

then, due to our selection of the function $a(t)$, we get

$$a(t_1) = \ldots = a(t_n) = m.$$

Thus, by spending the time

$$t_1 + \ldots + t_n = n \cdot 1 = n,$$

we get a uniformly m-successful (n, n)-learning strategy. For this strategy, the overall grade is equal to

$$F(m, \ldots, m) = m.$$

Let us now consider another (n, n)-learning strategy $(t'_, \ldots, t'_n)$, in which $t'_i = 1 - \varepsilon$ and

$$t'_1 = \ldots = t'_{i-1} = t'_{i+1} = \ldots = t'_n = 1 + \frac{\varepsilon}{n-1}.$$

In this case,

$$t'_1 + \ldots + t'_n = (t'_1 + \ldots + t'_{i-1} + t'_{i+1} + \ldots + t'_n) + t'_i =$$

$$(n-1) \cdot \left(1 + \frac{\varepsilon}{n-1}\right) + (1 - \varepsilon) =$$

$$(n-1) + \varepsilon + 1 - \varepsilon = n,$$

i.e., the same time as before: $t' = n$.

Under this learning strategy, the grades for different assignments are equal to $a(t'_i) = a(1-\varepsilon) = m - \varepsilon$ and

$$a(t'_1) = \ldots = a(t'_{i-1}) = a(t'_{i+1}) = \ldots = a(t'_n) =$$

$$a\left(1 + \frac{\varepsilon}{n-1}\right) = 1.$$

Thus, the overall grade is equal to

$$F(a(t'_1), \ldots, a(t'_{i-1}), a(t'_i), a(t'_{i+1}, \ldots, a(t'_n)) =$$

$$F(1, \ldots, 1 \ (i-1 \text{ times}), m-\varepsilon, 1 \ldots, 1).$$

Since $a(t'_i) = m - \varepsilon < m$, this (n, n)-learning strategy is *not* uniformly m-successful.

Let us consider the 2-element set

$$\mathcal{S} \stackrel{\text{def}}{=} \{(t_1, \ldots, t_n), (t'_1, \ldots, t'_n)\}.$$

This set contains a uniformly m-successful (n, n)-learning strategy $(t_1, \ldots, t_n)$. By definition of a grading scheme that encourages students to learn all the material, this means that every $(\mathcal{S}, F)$-optimal (n, n)-learning strategy is uniformly m-successful. Since we have shown that the learning strategy $(t'_1, \ldots, t'_n)$ is *not* uniformly m-successful, we can conclude that this strategy cannot be $(\mathcal{S}, F)$-optimal. Thus, the overall grade corresponding to the learning strategy $(t'_1, \ldots, t'_n)$ must be smaller than the overall grade corresponding to the strategy $(t_1, \ldots, t_n)$, i.e.:

$$F(1, \ldots, 1 \ (i-1 \text{ times}), m-\varepsilon, 1 \ldots, 1) < m.$$

The statement is proven.

2.3°. We have just proven that the inequality from Part 2.2 holds for every $\varepsilon > 0$. Since the grading scheme $F(a_1, \ldots, a_n)$ is a continuous function, in the limit $\varepsilon \to 0$, we get

$$F(1, \ldots, 1 \ (i-1 \text{ times}), m, 1 \ldots, 1) \le m.$$

2.4°. Let us now conclude the proof of Part 2, by proving that

$$F(a_1, \ldots, a_n) = \min(a_1, \ldots, a_n).$$

Indeed, let m denote the minimum $\min(a_1, \ldots, a_m)$, and let i denote the smallest value a_i, for which $a_i = m$ and $m \le a_j$ for all $j \ne i$.

2.4.1°. Since $a_j \le 1$, by monotonicity, we conclude that

$$F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n) =$$

$$F(1, \ldots, 1 \ (i-1 \text{ times}), a_i, 1 \ldots, 1) =$$

$$F(1, \ldots, 1 \ (i-1 \text{ times}), m, 1 \ldots, 1) \le m.$$

2.4.2°. Similarly, since $m = a_i \le a_j$ for all j, by monotonicity, we get

$$F(m, \ldots, m) \le F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n).$$

Since the function F is idempotent, we get

$$F(m, \ldots, m) = m$$

and hence,

$$m \le F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n).$$

2.4.3°. From

$$F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n) \le m$$

and

$$m \leq F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n),$$

we can now conclude that

$$F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n) = m,$$

i.e., that

$$F(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n) = \min(a_1, \ldots, a_n).$$

The theorem is proven.

VI. Resulting Recommendations Are Not That Unusual

The resulting recommendation is to take, as an overall grade for the class, the smallest of the grades gained for each module. At first, this may sound like a very radical idea, it is in line with what is usually done in grading.

For example, in our university, for a student to pass Calculus I, the student has to get satisfactory grade on each module. This corresponds to minimum. In some computer science classes, in order to pass a class, the student has to get a satisfactory grade both on the tests and on the labs.

Similarly, to get a degree, it is not sufficient for a student to have a good GPA, the student must get satisfactory grades on all required classes.

References

[1] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice-Hall, Upper Saddle River, New Jersey, 1995.

[2] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Boca Raton, Florida: Chapman & Hall/CRC, 2006.

[3] L. A. Zadeh, "Fuzzy sets", *Information and control*, vol. 8, pp. 338–353, 1965.