- ORIGINAL ARTICLE -

Formulation of an optimal academic exam
Formulación de un examen académico óptimo

Enrique E. Tarifa[1,2], Sergio L. Martínez[1], Samuel Franco Domínguez[1] and Jorgelina F. Argañaraz[2]

[1] *Faculty of Engineering, Universidad Nacional de Jujuy, San Salvador de Jujuy, Jujuy, 4600, Argentina*
{eetarifa, smartinez, sfrancodominguez}@fi.unju.edu.ar
[2] *CONICET, Argentina*
{eetarifa, jfarganaraz}@gmail.com

Abstract

The aim of this paper is to formulate an optimal academic exam for a given subject. To do this, the probability is first modelled of a student passing the exam according to the number of units he studies and the professor evaluates. That simulation model is developed by performing a probabilistic analysis. An optimal exam is then defined as the one that awards the grade that the student deserves. Therefore, in an optimal exam, approve those who deserve to approve, and disapprove those that do not deserve to approve. Besides, this exam must respect the limitations of time and effort that the professor imposes. Based on this definition and using the simulation model, an INLP type optimization model is formulated. This optimization model determines the number of units the professor must evaluate to maximize the probability of getting an optimal exam.

Keywords: Academic evaluation, optimization, probabilistic analysis.

Resumen

El objetivo de este trabajo es formular un examen académico óptimo para una materia dada. Para ello, primero, se modela la probabilidad de que un estudiante apruebe el examen en función del número de unidades que estudia y de las que el profesor evalúa. Ese modelo de simulación es desarrollado realizando un análisis probabilístico. Un examen

óptimo es luego definido como aquel que asigna la nota que el estudiante merece. Por lo tanto, en un examen óptimo, aprueban quienes merecen aprobar, y desaprueban quienes no merecen aprobar. Además, el examen debe respetar las limitaciones de tiempo y esfuerzo que el profesor impone. En base a esta definición y usando el modelo de simulación, se formula un modelo de optimización del tipo INLP. Este modelo de optimización determina el número de unidades que el profesor debe evaluar para maximizar la probabilidad de conseguir un examen óptimo.

Palabras claves: Análisis probabilístico, evaluación académica, optimización.

1. Introduction

This work is a substantially extended version of a previous one published at the conference CACIC 2017 [1]. This enhanced version contains more detailed explanations of models, additional results and a deeper analysis.

Evaluation is a critical issue in any institution, particularly in educational ones. According to Frola and Velásquez [2], the evaluation process involves information acquisition, elaboration of judgements once the information is processed, and the consequent decision-making aimed at improving processes and services.

Evaluation can be qualitative or quantitative. The first one is preferred for evaluating learning, while the second is chosen to measure the knowledge the student has retained at the end of a period.

Despite the importance of evaluation, it is not properly solved in the education field, with the consequent negative impact on the education of the students [3,4,5]. For example, Trillo Alonso and Porto Currás [6] have indeed analyzed the perception students had about evaluation in the Faculty of Educational Sciences of the University of Santiago de Compostela during the 1997-1998 academic year, and concluded that, for students,

evaluations did not achieve their objectives. Faced with this result, the authors concluded that, if this happened in a faculty of educational sciences, a better scenario was not likely to be met in other faculties. The results are even more worrying if it is considered that these students could be future professors.

Information and communication technologies (ICT) not only influence many aspects of education; but also influence evaluation particularly. ICT introduce new ways of evaluation, opening new possibilities by automating corrections, calculating statistical indices and performing histograms [7]. These functions favor a better evaluation; but they do not provide a finished solution; in fact, they bring about new problems. Multiple choice tests, for example, pose the problem of how to qualify those tests that have been made using random choices [8].

Huapaya *et al.* [9] stated that in order to carry out a fair evaluation, in addition to the grade of an exam, other aspects must be also considered; for example, averages of student's grades, class average, evolution of student's grades. These data were processed by a fuzzy logic expert system to diagnose the level of knowledge of the students. The use of an expert system has the advantage of removing subjectivity from the evaluation and producing uniform evaluations. It is however limited by the knowledge of the consulted experts and by the knowledge acquisition process [10].

An important aspect of evaluation that is little investigated is the professor role. When designing the evaluation, the professor makes several decisions: the number of questions to be asked, the topics covered by the questions, the approval grade (if it is not set by the institution). The decisions that the professor makes at this stage have a profound impact on the evaluation results. For this reason, the present work analyzes the effects that the decisions of the professor have on the evaluation.

To pose the problem formally, it is assumed that there is an evaluation stage in the lecturing of all subjects. This stage, in general, consists of a quantitative examination. The students pass the exam if they answer appropriately to the questions the professor asks. If the students have the minimum required level of knowledge, they obtain the minimum approval grade. In this scenario, the objective of this work is to optimize the evaluation. For this to be done, the probability of a student approving the evaluation must be first modeled as a function of a set of relevant variables.

The simulation model presented in this paper is developed after performing a probabilistic analysis. With the developed model, the work analyzes how the probability of the student passing the exam varies, and also analyzes the probabilistic distribution of the grades that can be obtained by the student depending on the selected variables.

Considering the previous analysis, an optimal exam is defined as the one that awards the grade that the student deserves. Therefore, in an optimally designed exam, approve those who deserve to approve, and disapprove those that do not deserve to approve. Besides, the exam should respect the limitations of time and effort that the professor imposes. Based on this definition and using the simulation model, an optimization model is formulated that determines the number of units that must be evaluated to maximize the probability of having an optimal exam. The optimal number of units to be evaluated produced by this model is not always the maximum allowed by the imposed restriction. In this sense, it is clear that sometimes fewer questions allow a better evaluation.

A second optimization model is then proposed that allows solving the problem within a certain tolerance. By increasing the tolerance, the number of units to be evaluated is significantly reduced. In turn, raising the approval threshold has little effect on the number of units to be evaluated. It is finally observed that increasing the maximum grade of the test significantly reduces the probability of assigning a fair grade to a student.

2. Simulation model

2.1. Problem formulation

The initial problem addressed in this paper is to estimate the probability that a student passes the exam of a subject composed of UM units of evaluation when the student studies UE units and the professor examines UT units. The exam is passed with a grade equal to or higher than NA, with a maximum grade equal to NM. The units of evaluation of the subject represent the degree of detail in which the professor breaks down the subject to carry out the evaluation. In degree of increasing detail, the units of evaluation can be units of the program of the subject, topics of the subject or subtopics of the subject. The answer of a student to an unit of evaluation can be correct or incorrect. Intermediate results are not considered. It is assumed that all units of evaluation have the same degree of difficulty, either for being studied or for being evaluated.

2.2. Presentation case

To best understand the solutions that will be presented in this work, it is convenient first to analyze a simple case. Table 1 contains all the possible exams for a subject with $NM = 10$, $NA = 4$, $UM = 5$ and $UT = 3$, where the "X" marks the units

the professor evaluates.

Table 1 Possible exams for $UM = 5$ and $UT = 3$.

Nº	1	2	3	4	5
1			X	X	X
2		X		X	X
3		X	X		X
4		X	X	X	
5	X			X	X
6	X		X		X
7	X		X	X	
8	X	X			X
9	X	X		X	
10	X	X	X		

If the student knows two units, $UE = 2$, and if it is assumed that they are the first two (generality is not lost with this assumption), the only favorable exams the student will pass are those that contain the two units he knows; *i.e.*, cases 8, 9 and 10. In those cases, the student will pass with a grade of 10 2/3 $\approx$ 7, with a probability equal to (favorable exams)/(possible exams) = 3/10.

2.3. Analytical solution

To present the analytical solution, it is convenient to analyze Table 2. In the first two rows, the content of the subject is divided into two parts: the UE units the student knows and the rest of the units of the subject, $UM-UE$. In the third row, the units evaluated in the exam are represented. Of the UT units asked by the professor in the exam, the student can only answer U units because they correspond to the part he studied, while the rest, $UT-U$, remain unanswered because they correspond to the part the student did not study.

Table 2 Exam structure.

..	U E	UE +1	UE +2	..	U M
UE			UM-UE		
U			UT-U		

With the structure presented in Table 2, the total number of possible exams is equal to the chosen UT combinations of UM, $C_{UM,UT}$. The number of exams in which the student can answer U questions can be calculated by considering that there are $C_{UE,U}$ possible combinations for the first part of the exam (the part the student knows), while there exist $C_{UM-UE,UT-U}$ possible combinations for the second part of the exam (that the student does not know). Therefore, the number of exams in which the student will answer U questions is $C_{UE,U} \, C_{UM-UE,UT-U}$. Hence, the probability of the student correctly answering u questions, $P(U = u)$, is given by the following expression:

$$Pu(u) = \frac{\binom{UE}{u}\binom{UM-UE}{UT-u}}{\binom{UM}{UT}} \quad \text{(Eq. 1)}$$

and the grade the student gets is:

$$N(u) = \text{Round}\left(NM \, \frac{u}{UT}\right) \quad \text{(Eq. 2)}$$

Continuing with the analysis of the structure shown in Table 2, it follows that the minimum value of U is:

$$U_{\min} = \max\left(0, UT - (UM - UE)\right) \quad \text{(Eq. 3)}$$

whereas the maximum value that U can reach is:

$$U_{\max} = \min(UE, UT) \quad \text{(Eq. 4)}$$

For the presentation case, the minimum and maximum values of U are 0 and 2, respectively.

The $Pu(u)$ distribution obtained is the "hypergeometric" one [10], a discrete distribution related to random sampling without replacement. For the problem that is being analyzed, there is a population of UM elements belonging to two categories: units the student knows and units he does not know; the EU units belong to the first, and the $UM-UE$ units belong to the second. Once these categories are defined, the hypergeometric distribution allows calculating the probability of obtaining elements of the first category in a sample without replacement of UT elements from the original population; *i.e.*, the probability that the student answers well U questions of the UT ones made by the professor.

From the probabilistic distribution of U, the probability PA can be derived, which is the probability that the student approves the evaluation by obtaining a grade $N(u)$ equal to or greater than NA:

$$PA = \sum_{u=U_{\min}}^{U_{\max}} H\left(N(u) - NA\right) Pu(u) \quad \text{(Eq. 5)}$$

$H(.)$ is the step function:

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \text{(Eq. 6)}$$

On the other hand, the probability of the student obtaining a grade n in the exam, $P(N = n)$, is calculated as follows:

$$PN(n) =$$

$$\sum_{u=U_{\min}}^{U_{\max}} H\left(N(u) - n\right) H\left(n - N(u)\right) Pu(u) \quad \text{(Eq. 7)}$$

i.e., the probability of n is obtained by adding the probabilities of all u with $N(u) = n$.

For the presentation case, $PA = Pu(2) = 3/10$, with $N(2) = 7$. In this calculation, only the probability of $u = 2$ remains because $N(0) = 0$ and $N(1) = 3$.

Finally, the value of the fair grade NJ can be

defined. This is the grade deserved by the student who studied *UE* units:

$$NJ = \text{Round}\left(NM\,\frac{UE}{UM} \right) \quad \text{(Eq. 8)}$$

If this grade is equal to or higher than *NA*, the student deserves to approve the exam; otherwise he does not deserve to approve. For the presentation case, *NJ* = 4, but the only possible grades are 0, 3 and 7. It can be seen that is not always possible to qualify with the fair grade.

3. Case study

The results obtained for a subject with *NM* = 10, *NA* = 4 and *UM* = 10 are presented below. In this case, *NJ* will be equal to *UE*, so 4 is the minimum amount of units that the student must study to deserve passing the exam. Although the graphics to be presented in this work should only have points, lines were included to ease the recognition of the different series.

Fig. 1 shows the probability *PA* of a student passing the exam when he studies *UE* units and the professor asks *UT* units. In this figure, as *UT* increases, the exam better discriminates among students who deserve to pass (*UE* ≥ *NA*) and those who do not. For *UT* equal to 3, 6 and 9, *PA* decreases. To explain this behavior, the case with *UE* = 3 is analyzed. On the one hand, if *UT* = 2, *u* may be 0, 1 or 2; with grades 0, 5 or 10, respectively; therefore, *PA* = 0.47 + 0.07 = 0.54. On the other hand, if *UT* = 3, *u* may be 0, 1, 2 or 3; with grades 0, 3, 7 or 10, respectively; therefore, *PA* = 0.18 + 0.01 = 0.19. Hence, the main reason for the decrease of *PA* is that, in the last case, there is an additional exam that is not passed by the student, in which he answers only one question correctly (*N*(1) = 3 < *NA*); whereas for the same situation in the first case, since there is one less question, the student passes the exam (*N*(1) = 5 > *NA*).

Fig. 2 has the same information than Fig. 1, but from another point of view. From this new perspective, it is more clearly seen that, as *UT* increases, the approval threshold is *NA*. Fig. 1 is more useful for the professor, since it shows the effect of *UT* on *PA*. Fig. 2 instead, is more useful for the student, since it shows the effect of *UE* on *PA*.

Fig. 3 to Fig. 11 show the probability distributions *PN*(*n*) of the grades a student may obtain for different amounts of studied units *UE*. In those plots, the number of units *UT* the professor evaluates is a parameter. As it can be seen, the probable grades may be quite different from the fair grade; *e.g.,* for *UE* = 5 (*NJ* would also be 5) and *UT* = 2 (Fig. 7), the probable grades are 0, 5 and 10. In other words, a student that deserves a grade equal

to 5, may get 0, 5 or 10. Besides, those figures show that the grade dispersion decreases when *UT* increases. In contrast, the grade dispersion is almost independent of *EU*. In conclusion, only the professor can reduce the grade uncertainty.
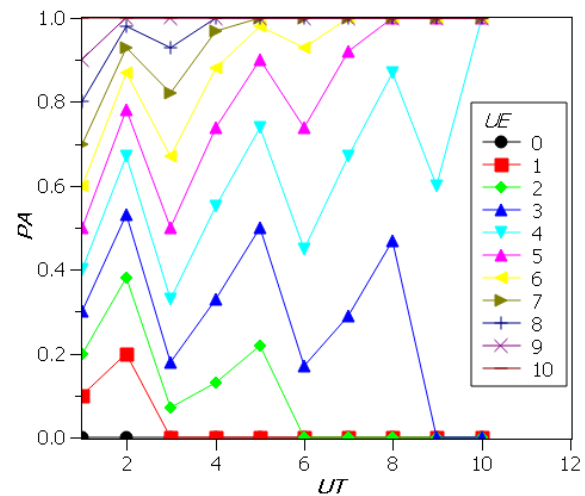

Fig. 1 Probability *PA* of passing the exam when the student studies *UE* units and the professor asks *UT* units (*UE* as parameter), with *NM* = 10, *NA* = 4 and *UM* = 10.
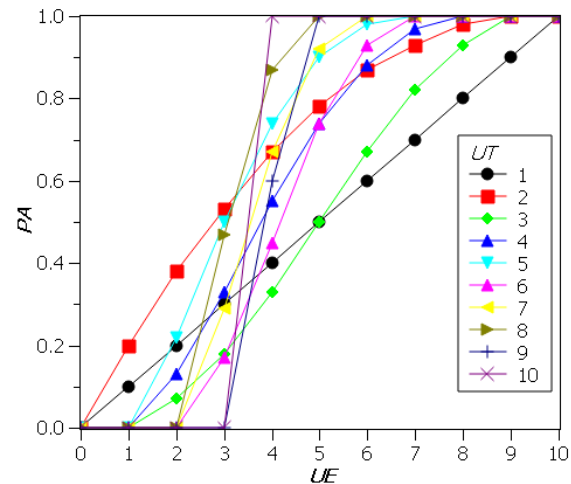

Fig. 2 Probability *PA* of passing the exam when the student studies *UE* units and the professor asks *UT* units (*UT* as parameter), with *NM* = 10, *NA* = 4 and *UM* = 10.

Fig. 3 Probabilistic distribution of the grades *Pn*(*n*) for *UE* = 1, with *NM* = 10, *NA* = 4 and *UM* = 10.
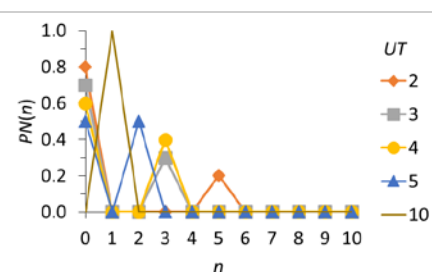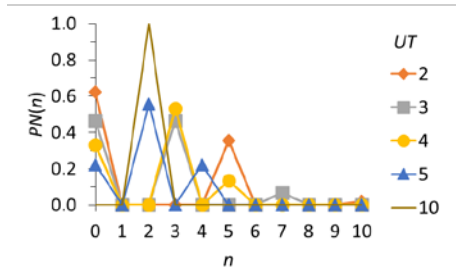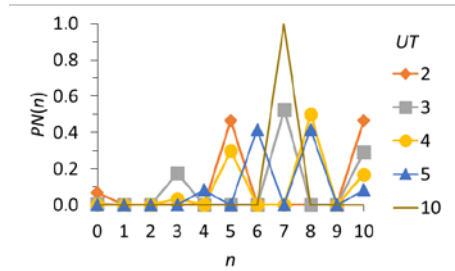
Fig. 4 Probabilistic distribution of the grades $Pn(n)$ for $UE = 2$, with $NM = 10$, $NA = 4$ and $UM = 10$.
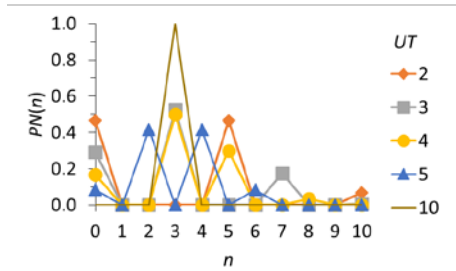

Fig. 5 Probabilistic distribution of the grades $Pn(n)$ for $UE = 3$, with $NM = 10$, $NA = 4$ and $UM = 10$.
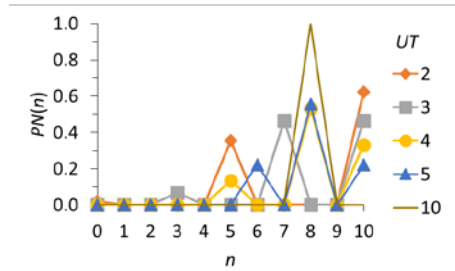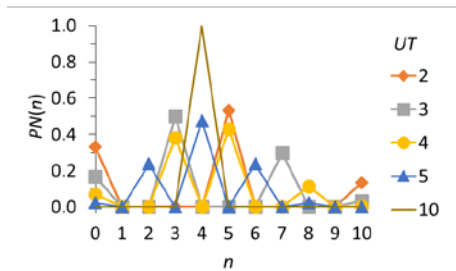

Fig. 6 Probabilistic distribution of the grades $Pn(n)$ for $UE = 4$, with $NM = 10$, $NA = 4$ and $UM = 10$.

Fig. 7 Probabilistic distribution of the grades $Pn(n)$ for $UE = 5$, with $NM = 10$, $NA = 4$ and $UM = 10$.

Fig. 8 Probabilistic distribution of the grades $Pn(n)$ for $UE = 6$, with $NM = 10$, $NA = 4$ and $UM = 10$.
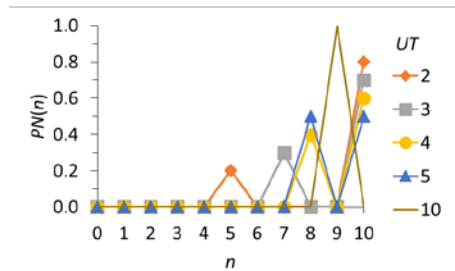

Fig. 9 Probabilistic distribution of the grades $Pn(n)$ for $UE = 7$, with $NM = 10$, $NA = 4$ and $UM = 10$.

Fig. 10 Probabilistic distribution of the grades $Pn(n)$ for $UE = 8$, with $NM = 10$, $NA = 4$ and $UM = 10$.

Fig. 11 Probabilistic distribution of the grades $Pn(n)$ for $UE = 9$, with $NM = 10$, $NA = 4$ and $UM = 10$.
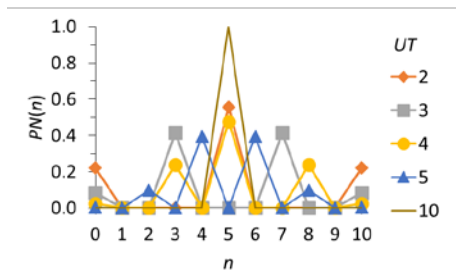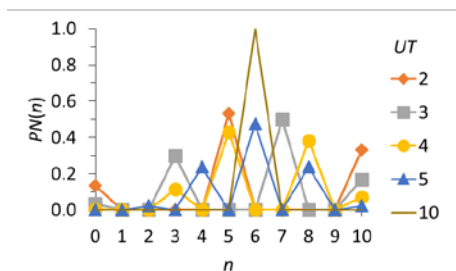
4. The optimal exam

As it was stated in the introduction of the present work, an optimal exam is defined as the one that awards the grade that the student deserves, and which can be carried out respecting the time and effort limitations that the professor imposes (time of completion of the exam, time of correction, etc.). Consequently, in an optimal examination, approve those who deserve to approve, and disapprove those that do not deserve to approve. Thus, for a student who studied UE units, the following objective functions to be maximized can be proposed:

- Probability of fair grade

$$f_1(UE) = PN(NJ) \quad \text{(Eq. 9)}$$

This function aims at giving the student the grade he deserves. This is a difficult measure to satisfy because the student must be assigned to one of the $NM + 1$ possible categories (by assigning a grade from 0 to NM); and this is increasingly difficult as NM becomes higher.

- Probability of fair approval

$$f_2(UE) = \begin{cases} 1 - PA & NJ < 4 \\ PA & NJ \geq 4 \end{cases} \quad \text{(Eq. 10)}$$

This function aims at approving if the student deserves it, and disapproving if he does not deserve to pass. This is a more relaxed measure than the previous one because it considers only two categories: approved and disapproved.

The posed objective functions measure how optimal the exam is in relation to a particular type of student: the one who studied UE units. To consider all students, a global objective function independent of UE must be considered. Two possible global objective functions that meet this condition are:

$$FO_1(UT) = \sum_{ue=0}^{UM} Pb(ue) f_1(ue) \quad \text{(Eq. 11)}$$

$$FO_2(UT) = \sum_{ue=0}^{UM} Pb(ue) f_2(ue) \quad \text{(Eq. 12)}$$

$Pb(ue)$ is the probability that a student studies ue units, $P(UE = ue)$. Therefore, the proposed global objective functions are probabilistic averages of the individual objective functions analyzed before.

Particularly in this work, it is assumed that $Pb(ue)$ obeys a binomial distribution:

$$Pb(ue) = \binom{UM}{ue} p^{ue} (1-p)^{UM-ue} \quad \text{(Eq. 13)}$$

where p is the probability that students study a given unit. The higher p, the more dedicated the students are (Fig. 12). For the case study, a value of $p = 0.5$ is adopted. Fig. 13 shows how the two global objective functions vary for that case.

The presented global objective functions must be maximized to achieve an optimal exam. The only decision variable the professor has is UT. In Fig. 13, both functions reach the maximum value when $UT = UM$. However, this solution has the highest cost (duration of the exam design, duration of the exam, duration of the correction). Therefore, it is convenient to establish the following constrain:

$$UT \leq UT_{\max} \quad \text{(Eq. 14)}$$

for preventing a high cost. $UT_{\max}$ is the maximum number of units the professor can or wants to evaluate. For the case study, $UT_{\max} = UM/2 = 5$.

Fig. 3 to Fig. 11 show, for the case study ($NM = 10$, $NA = 4$, $UM = 10$, $p = 0.5$ and $UT_{\max} = 5$), four low-cost solutions that roughly meet the first proposed objective function. These solutions are obtained for UT from 2 to 5. If the lines corresponding to those UT values are examined in Fig. 2, it can be seen that the option $UT = 3$ tends to disapprove in general (even those that deserve to approve); while the option $UT = 5$ tends to approve in general (even those students who do not deserve to approve). $UT = 4$ is an intermediate option, and apparently the most recommended; but the effect of p has not yet been considered. To do this, Fig. 13 is

analyzed, which is constructed for $p = 0.5$. In this figure, it can be seen that the trivial solution $UT = 5$ satisfies the constraint stated above (less than or equal to 5) and maximizes both objective functions. Then, for the case study, it is advisable to evaluate five units. It is important to note however that, although both objective functions have an increasing tendency, they are not monotonically increasing functions. Hence, in some cases, there may exist non-trivial solutions to the proposed optimization problem.

Finally, it can also be seen that being the second objective function more relaxed than the first one, the second function has higher values and is less sensitive to UT. This means that an exam is more likely to be considered optimal if it is evaluated with the second function.
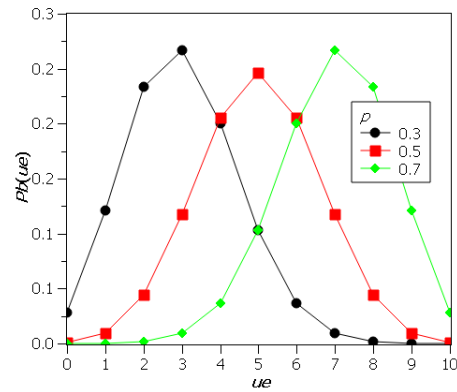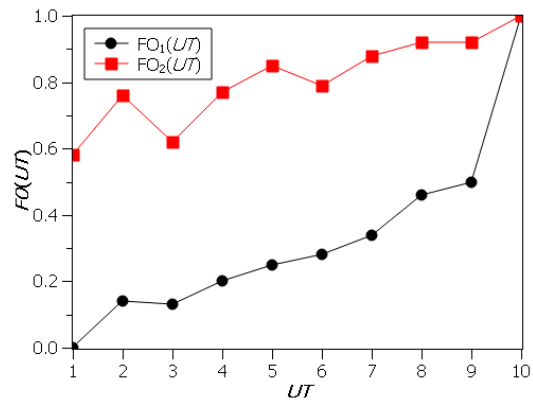

Fig. 12 Binomial distribution of UE.

Fig. 13 Global objective functions for the case study, with $NM = 10$, $NA = 4$, $UM = 10$ and $p = 0.5$.

4.1. Optimization model

To solve the stated problem for any UM and p, the following INLP (Integer Nonlinear Programming) optimization model is posed:

$$\begin{aligned} &\underset{UT}{\text{Max}} \; FO(UT) \\ &\text{s. t.:} \\ &UT \leq UT_{\max} \\ &UT \in \mathbb{Z} \end{aligned} \quad \text{(Eq. 15)}$$

where $FO(UT)$ is the chosen global objective function, $FO_1(UT)$ or $FO_2(UT)$. Because the feasible region is small, this model can be solved by exhaustive search.

The solution produced by this model guarantees the maximum value of $FO(UT)$, but not the minimum value of UT. In the case of multiple solutions, the desired solution is the minimum UT. To find this value, a second INLP optimization problem must be solved:

$$\underset{UT}{\text{Min}} \, UT$$

$$\text{s. t.:}$$

$$FO(UT) \geq FO_{\text{opt}} \qquad \text{(Eq. 16)}$$

$$UT \in \square$$

where FO_{opt} is the maximum value of the objective function reported by the first optimization problem.

Table 3 Optimum values of UT for FO_1 and FO_2.

p		0.3		0.5		0.7	
UM				UT			
	max	1opt	2opt	1opt	2opt	1opt	2opt
5	2	*1*	2	*1*	2	*1*	2
6	3	3	3	3	*2*	3	*2*
7	3	3	3	3	*2*	3	*2*
8	4	4	4	4	4	4	4
9	4	*3*	*3*	*3*	4	4	4
10	5	5	*4*	5	5	5	5
11	5	5	5	*4*	5	5	5
12	6	6	6	6	*5*	6	*5*
13	6	6	6	6	*5*	6	*5*
14	7	7	7	7	*5*	7	7
15	7	7	*6*	*6*	7	7	7
16	8	8	*7*	8	8	8	8
17	8	*5*	8	8	8	8	8
18	9	9	9	*8*	*8*	9	*8*
19	9	9	9	*8*	*8*	9	*8*
20	10	10	10	10	*8*	10	10
21	10	10	*9*	10	*8*	10	10
22	11	11	*10*	11	11	11	11
23	11	*10*	*9*	11	11	*10*	11
24	12	12	12	*11*	*11*	12	*11*
25	12	12	12	*11*	*11*	12	*11*
26	13	*12*	*12*	13	*11*	13	13
27	13	*12*	*12*	13	*11*	13	13
28	14	*12*	*12*	14	14	*13*	14
29	14	*12*	*12*	14	14	*13*	14
30	15	15	15	15	*14*	15	*14*

Table 3 shows the optimal UT values obtained

for different values of UM, UM_{max} and p. The $UT1_{\text{opt}}$ values were obtained with $FO(UT) = FO_1(UT)$; while the $UT2_{\text{opt}}$, with $FO(UT) = FO_2(UT)$. Fig. 14 shows the results corresponding to $p = 0.5$. It can be seen that the quantity UT does not always reach the maximum value allowed by the constraint (up to 3 units less than the maximum amount allowed), and hence sometimes fewer questions produce a better evaluation. These somewhat unexpected cases are marked with red and italic font in the table. Additionally, as the second objective function is more relaxed than the first one, the optimal values of $UT2_{\text{opt}}$ are less or equal to $UT1_{\text{opt}}$, producing lower evaluation costs.
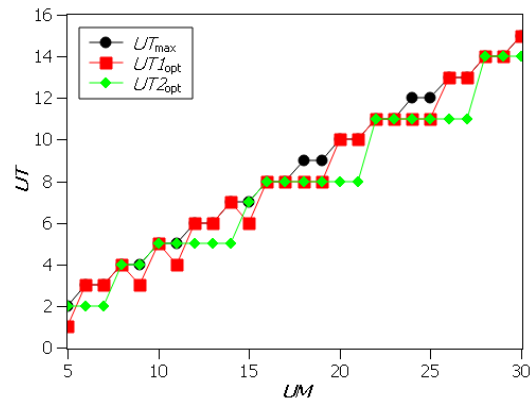

Fig. 14 Optimal values of UT for FO_1 and FO_2, with $NM = 10$, $NA = 4$ and $p = 0.5$.

4.2. Optimization model with tolerance

Inspection of Fig. 13 hints that there is no great difference between $FO_1(2)$ and $FO_1(5)$, nor between $FO_2(2)$ and $FO_2(5)$. For this reason, if the professor has a certain tolerance, the practical amount of units to evaluate, UT_{pra}, may be less than the UT_{opt} recommended in the previous section.

With this new tolerance parameter, the optimization problem gets broken down into two sequential problems. Firstly, the first optimization problem posed in the previous section must be solved, in order to determine the maximum value FO_{opt} of the objective function. With that value, the following INLP optimization problem must then be solved:

$$\underset{UT}{\text{Min}} \, UT$$

$$\text{s. t.:}$$

$$FO(UT) \geq FO_{\text{opt}} - Tol \qquad \text{(Eq. 17)}$$

$$UT \in \square$$

where $Tol \in [0, 1]$ is the tolerance, or decrease in the probability of performing an optimal examination accepted by the professor.

Applying this model to the case study ($NM = 10$, $NA = 4$, $UM = 10$, $p = 0.5$ and $UT_{\text{max}} = 5$) with $Tol = 0.1$, the practical amount of units to be evaluated is 2 in place of 5 for the second global

objective function. This is a significant reduction in the cost of the evaluation.

Fig. 15 and Fig. 16 present the results for *Tol* values equal to 0.1 and 0.2, respectively, with $p = 0.5$ and the same UT_{max} considered in Table 3. The $UT1_{pra}$ values were obtained with $FO(UT) = FO_1(UT)$, and $UT2_{pra}$ were obtained with $FO(UT) = FO_2(UT)$. In these figures, the UT values are well below the allowed amounts UT_{max} and the optimum values UT_{opt}. This is achieved without affecting too much the quality of the evaluation (*i.e.*, $FO(UT)$).
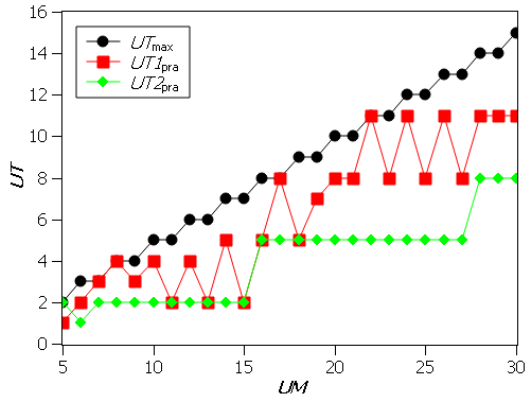

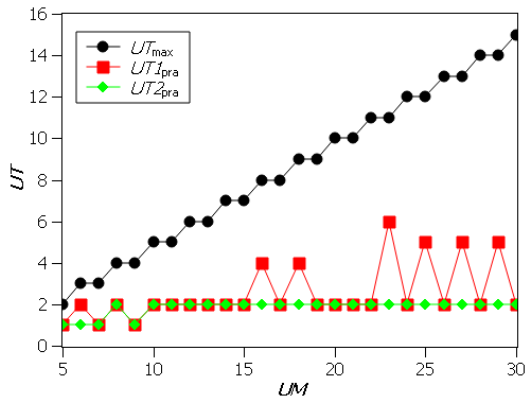Fig. 15 Optimal values of UT for FO_1 and FO_2, with $NM = 10$, $NA = 4$, $p = 0.5$ and $Tol = 0.1$.

Fig. 16 Optimal values of UT for FO_1 and FO_2, with $NM = 10$, $NA = 4$, $p = 0.5$ and $Tol = 0.2$.

4.3. Approval grade effect

In the previous study, it was assumed that the approval grade NA was equal to 4, which is the standard for approval in some universities. However, other universities or chairs adopt different approval grades according to the instance of the evaluation. For example, a chair may adopt a grade 5 to approve a partial exam, a grade 7 to promote the subject and a grade 4 to approve the final exam. Considering this possible scenario, the effect of modifying NA should be analyzed.
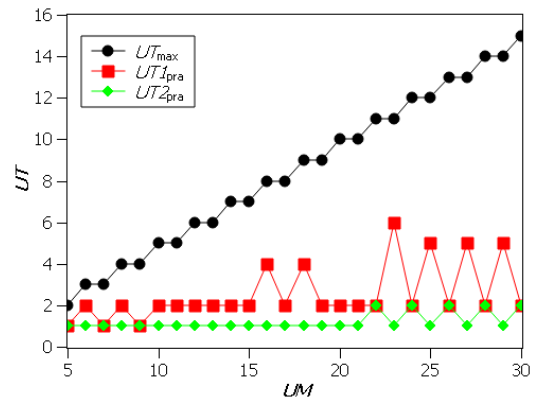

Fig. 17 Optimal values of UT for FO_1 and FO_2, with $NM = 10$, $NA = 5$, $p = 0.5$, $UT_{max} = 5$ and $Tol = 0.2$.

The modification of the approval grade modifies $FO_2(UT)$ and, therefore, also $UT2_{opt}$ and $UT2_{pra}$. Fig. 17 shows the values of $UT2_{pra}$ for $NM = 10$, $NA = 5$, $UM = 10$, $p = 0.5$, $UT_{max} = 5$ and $Tol = 0.2$. The same values are obtained for $NA = 6$. The values obtained for $NA = 7$ are the same as those obtained for $NA = 4$ (Fig. 16). Although there are changes in $UT2_{pra}$ for $Tol = 0.2$ when NA changes from 4 to 7, the difference is at most one unit of evaluation.

4.4. Scale effect

When the maximum grade NM is increased to 100 and NA to 40, a marked deterioration of $FO_1(UT)$ is observed, whereas $FO_2(UT)$ remains almost unchanged. Fig. 18 presents both global objective functions for $NM = 100$, $NA = 40$, $UM = 10$ and $p = 0.5$. $FO_1(UT)$ is degraded because when the scale is increased the possible grades also increase, and then the exam grade is less likely to match the fair grade NJ. In this case, both $UT1_{pra}$ and $UT2_{pra}$ adopt values less than or equal to 2 when $UT_{max} = 5$ and $Tol = 0.2$ (Fig. 19).
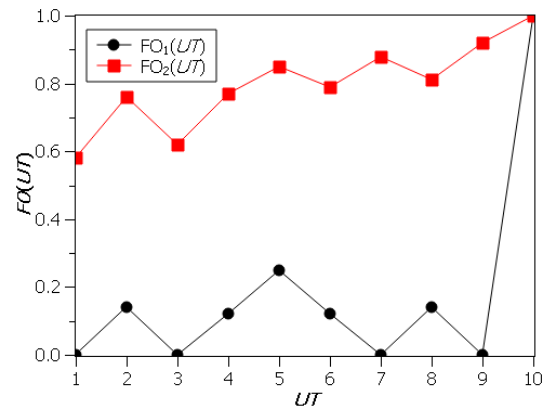

Fig. 18 Global objective functions, with $NM = 100$, $NA = 40$, $UM = 10$ and $p = 0.5$.
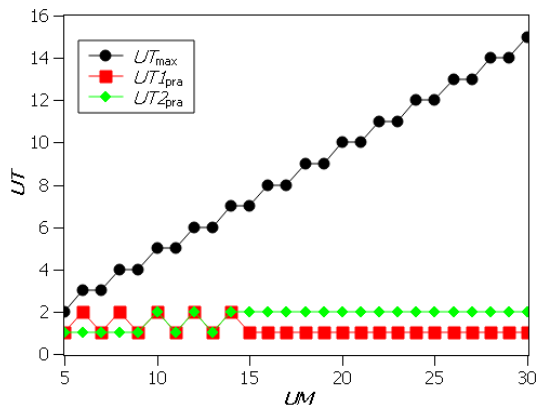
Fig. 19 Practical values of UT for FO_1 and FO_2, with $NM = 100$, $NA = 40$, $p = 0.5$, $UT_{max} = 5$ and $Tol = 0.2$.

4.5. Grade error

The effect of the grade error that the professor tolerates should also be analyzed. In this case, the professor accepts as good a grade belonging to the interval $NJ \pm Error$. The simplest way to deal with this case is to change NM in the following way:

$$NM' = \left\lceil \frac{NM}{2\ Error} \right\rceil \qquad \text{(Eq. 18)}$$

where NM' is the new scale. This solution is not exact, but it is a good approximation.

To see how good this approximation can be, the case in which the professor has a scale with $NM = 100$ and $Error = 5$ is analyzed. For this case, NM' will be equal to 10. Fig. 20 shows the lower limits (LI and LI') and upper limits (LS and LS') of the grades that the professor will assign with both scales. The abscissae of the figure represent the grades the student deserves. The ordinates represent the limits of the grades assigned by the professor. It can be seen, that there is little error in replacing the band corresponding to $NM = 100$ with the band corresponding to $NM' = 10$.
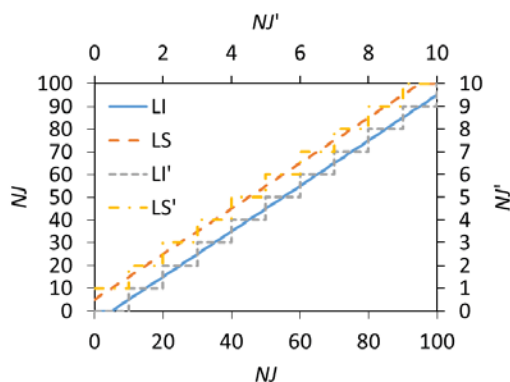

Fig. 20 Upper and lower limits of grades.

4.6. Sequence of questions

If the exam is oral, the following procedure can be implemented to accelerate the evaluation:

1. The UT units of evaluation are presented to the student.
2. The student is allowed to choose the answering order after being advised to choose an order of his convenience. The student should put units he dominates the most in the first places and leave units he scarcely know in the last ones.
3. The units are examined following the sequence chosen by the student.
4. The exam is ended when the student correctly presented all the units or when the student could not present a unit.
5. The exam grade is calculated by assigning to U the number of units the student correctly presented.

It should be noted that this procedure does not alter the probability of assigning the fair grade, nor the probability of approving for those who deserved it, nor the probability of disapproving for those who should not pass the exam. The only consequence of this procedure is the reduction of the duration of the exam. If one student cannot answer correctly an evaluation unit, he will not be able to answer the following questions because, by ordering the questions, the student acknowledged that knows even less of the following topics.

4.7. Subunits evaluation

Another common case is that in which the professor divides the evaluation units into subunits. That is, once the UT units for the examination have been chosen, each unit is further divided into subunits for evaluation. To model this situation, the student should also be allowed to decompose into subunits the UE chosen units. With this variation one restriction must be removed, the one related to the student answering correctly or not at all each unit of evaluation. Now this must be applied to each subunit separately. This subdivision does not modify the probable grades determined with the previous study, but it adds intermediate grades: the greater the number of subunits in which the units are decomposed, the greater the number of feasible intermediate grades.

5. Conclusions

In this paper, a simulation model was presented to estimate the probability of a student passing an exam. The probabilistic distribution of the grades to be obtained was also analyzed. Based on a previous analysis of the examination practice, an optimal exam was defined as the exam that awards the grade that the student deserves, and that can be carried out with respect to the limitations of time and effort of

the professor. Briefly, an optimal exam approves those who deserve to approve and disapproves those that do not deserve to approve. Based on this definition, an optimization model was formulated that determines the number of units to be evaluated in order to maximize the probability of carry out an optimal exam. This model was solved for different cases, and it was found that the optimal number of units to be evaluated was not always the maximum allowed.

A second optimization model was then proposed that allowed solving the evaluation problem with a certain tolerance. By increasing the tolerance accepted by the professor, the number of units to be evaluated was significantly reduced. The effects of modifying the minimum grade for approval were also analyzed. It was found that it does not have an important effect on the number of units that must be evaluated to get an optimal examination. On the other hand, modifying the maximum grade of the exam greatly reduced the probability of assigning a fair grade to the student.

The main practical teachings of this study are: (i) sometimes fewer questions produce a better evaluation; (ii) getting an optimal exam is more likely when the exam is only qualified as approved or disapproved (no grade scale).

Acknowledgements

Competing interests

The authors have declared that no competing interests exist.

List of symbols

Error: Error of a grade compared with *NJ*.
$f_1(UE)$: Probability of fair grade.
$f_2(UE)$: Probability of fair approval.
$FO(UT)$: The chosen global objective function.
$FO_1(UT)$: Average probability of fair grade.
$FO_2(UT)$: Average probability of fair approval.
FO_{opt}: The maximum value of $FO(UT)$.
LI: Lower limit for grades with scale *NM*.
LI': Lower limit for grades with scale *NM'*.
LS: Upper limit for grades with scale *NM*.
LS': Upper limit for grades with scale *NM'*.
$N(u)$: Exam grade of a student correctly answering *u* questions, $U = u$.
NA: Minimum grade to pass the exam.
NJ: Fair grade in the scale *NM*.

NJ': Fair grade in the scale *NM'*.
NM: Maximum of the scale grade.
NM': Maximum of the new scale grade.
p: Probability that students study a given unit.
PA: Probability of a student passing the exam.
$PN(n)$: Probability of a student getting a particular grade, $P(N = n)$.
$Pu(u)$: Probability of a student correctly answering *u* questions, $P(U = u)$.
Tol: Tolerance of performing an optimal examination accepted by the professor.
U: Number of right answers of a student.
UE: Number of units a student studies.
UM: Number of units of evaluation.
U_{max}: Maximum value of *U*.
U_{min}: Minimum value of *U*.
UT: Number of units the professor examines.
$UT1_{opt}$: Optimum value of *UT* when $FO(UT) = FO_1(UT)$ and $Tol = 0$.
$UT1_{pra}$: Optimum value of *UT* when $FO(UT) = FO_1(UT)$ and $Tol > 0$.
$UT2_{opt}$: Optimum value of *UT* when $FO(UT) = FO_2(UT)$ and $Tol = 0$.
$UT2_{pra}$: Optimum value of *UT* when $FO(UT) = FO_2(UT)$ and $Tol > 0$.
UT_{max}: Maximum value of *UT*.
UT_{opt}: Optimum value of *UT*, for which $FO(UT_{opt}) = FO_{opt}$.

References

[1] E. Tarifa, S. Martínez, S. Franco Domínguez and J. Argañaraz, "Formulación de un Examen Óptimo," in *XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017)*, pp. 322–331, 2017.

[2] P. Frola and J. Velásquez, *Competencias docentes para la evaluación cuantitativa del aprendizaje*. D. F. México: Centro de Investigación Educativa y Capacitación Institucional, 2011.

[3] J. Tata, "Grade Distributions, Grading Procedures, and Students' Evaluations of Instructors: A Justice Perspective", *The Journal of Psychology*, vol. 133, no. 3, pp. 263–271, 1999.

[4] D. Close, "Fair Grades," *Teaching Philosophy*, vol. 32, no. 4, pp. 361–398, 2009.

[5] J. Jones Miller, "A Better Grading System: Standards-Based, Student-Centered Assessment", *English Journal*, vol. 103, no. 1, pp. 111–118, 2013.

[6] F. Trillo Alonso and M. Porto Currás, "La percepción de los estudiantes sobre su evaluación en la universidad. Un estudio en la facultad de ciencias de la educación," *Innovación Educativa*, no. 9, pp. 55–75, 1999.

[7] J. Álamo Serrano, "Nuevas posibilidades de evaluación usando las TIC's: un vistazo a cuatro

casos," in *La evaluación de los estudiantes en la Educación Superior*, Valencia: Universidad de Valencia, 2018, pp. 54–73.

[8] J. González-Santander and G. Martín, "Análisis de la fórmula para la calificación de pruebas tipo test multi-respuesta," *Nereis. Revista Iberoamericana Interdisciplinar de Métodos*, no. 3, pp. 53–59, 2010.

[9] C. Huapaya, F. Lizarralde and G. Arona, "Modelo basado en Lógica Difusa para el Diagnóstico Cognitivo del

Estudiante," *Formación universitaria*, vol. 5, no. 1, pp. 13–20, 2012.

[10] G. Bojadziev, *Fuzzy Sets, Fuzzy Logic, Applications (Series on Advances in Mathematics for Applied Sciences)*. New York: World Scientific Publishing Company, 1996.

[11] C. Walck, *Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists*. Stockholm: University of Stockholm, 2007.