Isla-Díaz, Rosa; Marrero-Hernández, Hipólito; Hess-Medler, Stephany; Soriano, Mabel; Acosta-Rodríguez, Severo; Pérez-Monteverde, María-Victoria, & Blanco-Freijo, Marcos (2018). A longitudinal look: Is "Docentia" an useful instrument for university teacher's evaluation? *RELIEVE, 24*(2), art. 2. doi http://doi.org/10.7203/relieve.24.2.12142

Revista ELectrónica de Investigación y EValuación Educativa

RELIEVE

e-Journal of Educational Research, Assessment and Evaluation

ISSN: 1134-4032

Una mirada longitudinal: ¿Es el "Docentia" útil para la evaluación del profesorado universitario?

A Longitudinal Look: Is "Docentia" an useful instrument for university teacher's evaluation?

Isla-Díaz, Rosa; Marrero-Hernández, Hipólito; Hess-Medler, Stephany; Soriano, Mabel; Acosta-Rodríguez, Severo; Pérez-Monteverde, María-Victoria, & Blanco-Freijo, Marcos

Universidad de La Laguna (España).

Abstract

Docentia is an evaluation model proposed by ANECA (National Agency for Quality Evaluation) and used by Spanish universities to evaluate teaching quality. One of the detected problems is the low capacity to discriminate between the possible categories of teachers, showing a bias toward the highest score (Excellent). This affects negatively the prestige of the model: if most teachers are labelled "excellent", those teachers who really foreground above the set would not be detected. Thus, the capacity to orientate teachers to improve their performance will be minored. This study explores the discrimination capacity of the model based on the experience at the University of La Laguna (Docentia-ULL). We simulate results of the evaluation changing the model in two ways. On one hand, changing weights of some dimensions and sub dimensions, as well as maximums of some criteria, in order to increase the value of the commitment of teachers with training and educational innovation, opposite to the mere fulfilment of teaching obligations. On the other hand, increasing the required minimum for student's satisfaction and academic supervisor's satisfaction to become necessary requirement to obtain positive, very positive or excellent evaluation. Our results show that reducing the weight of teaching obligations and increasing the weight of training and innovation produce a more centered distribution. Most teachers were situated in the middle categories. We also compared the real results obtained after the implementation of the chosen alternative simulated model. We discuss the implications of those improvements on the evaluation of teaching quality and on the performance of Spanish university teachers.

Keywords: Docentia, Performance evaluation in higher education, university teaching, university teacher, student satisfaction in higher education.

Reception Date
2018 March 14

Approval Date
2018 October 1

Publication Date:
2018 October 9

Resumen

El Docentia es el modelo propuesto por ANECA, y asumido por las universidades españolas para la evaluación de la calidad docente. Uno de los problemas del modelo que se viene detectando es su escasa capacidad para diferenciar al profesorado entre las distintas categorías posibles, con un sesgo muy acentuado hacia la calificación de "Excelente". Esto afecta al prestigio del modelo: primero, si la gran mayoría del profesorado es "Excelente", no se verá reflejado el profesorado que realmente destaca sobre el conjunto y, segundo, afecta a su capacidad para orientar al profesorado hacia la mejora de su docencia. En el presente estudio examinamos la capacidad discriminativa del modelo en base a la experiencia de su implementación en la Universidad de La Laguna (Docentia-ULL). Hemos simulado el resultado de la evaluación cambiando el modelo en dos sentidos. Por un lado, los pesos de las dimensiones y sub-dimensiones, así como de los topes de cada criterio, para primar la valoración del compromiso del profesorado con la formación y la innovación educativa, frente al mero cumplimiento de las obligaciones docentes. Por otro lado, los resultados de satisfacción del alumnado y de los responsables académicos pasan a actuar como requisito independiente de los méritos del docente en la evaluación. Además, se ha elevado el mínimo necesario de estas

Fecha de recepción
2018 Marzo 14

Fecha de aprobación
2018 Octubre 1

Fecha de publicación
2018 Octubre 9

dimensiones para obtener una valoración "Favorable", "Muy Favorable" o "Excelente". Los resultados mostraron que la disminución del peso en la evaluación de las obligaciones docentes frente a la formación e innovación produjo una distribución de la calificación del profesorado más centrada hacia las categorías intermedias. Además se compara con resultados reales obtenidos después de la implantación del modelo alternativo elegido a raíz de las simulaciones. Se discuten las implicaciones de este tipo de mejoras en la evaluación de la calidad docente y del desempeño del profesorado en la universidad española.

Palabras clave: Docentia, Evaluación del desempeño en educación superior, docencia universitaria, profesorado universitario, satisfacción de estudiantes universitarios.

The "Docentia" as a model for teacher evaluation: background

The homologable evaluation of teachers is a requirement for universities within the European Higher Education Area (EHEA). In Spain, the so-called Docentia programme has been the reference model developed by the National Agency for Quality Assessment and Accreditation (ANECA) to evaluate university teaching staff. Based on this model and relying on common objectives such as the continuous improvement of training programmes or completing the quality assurance systems for the accreditation of degrees, Universities and Regional Assessment Agencies have designed their own programmes, with specific characteristics agreed upon with their university community, fulfilling, nevertheless, a common mandate: to contribute to the improvement of teaching quality, thus guaranteeing the quality of its teaching staff in accordance with the principles of the EHEA (Murillo, 2008).

The evaluation of teaching activity, traditionally approached through opinion surveys fulfilled by students (Calderón & Escalera, 2008), is not an easy task, due to the diversity of functions performed by university teaching staff, not always observable and therefore measurable through a system of effective performance evaluation. The university professor performs a complex set of tasks that can be grouped into three main types of functions: teaching, research and academic management (Caballero & Bolívar, 2015; Murillo, 2008;). This workload is difficult to evaluate objectively due to its intangible nature. Its evaluation, until now, is limited to the results obtained with each functional group. However, an appropriate evaluation system would have to take into account this complex functional diversity, particularly in the case of teaching activity (Álvarez Rojo et al., 2009), including specific measurement indicators for each group of task. In addition, it has to comply with the terms of an equation that has been found difficult to implement, because the involved functions are sometimes opposing rather than complementary due to that the work overload experienced by these professionals would cause that the adequate performance of a group of functions entail the abandonment of others.

Specifically, in the evaluation of the quality of teaching, the tasks to be taken into account for an efficient performance are even more diverse and complex, if possible, given the shift experienced by teaching as a consequence of pursuing compliance with EHEA university regulations. The aim of university teaching would be no longer the acquisition of knowledge by students, but rather the acquisition of competences through the student autonomous work, which entails different demands on the teaching staff. Related fundamentally, on the one hand, to vertical and horizontal planning and coordination, while, on the other hand, related to academic tutoring, but also to new, or not so new, requirements such as continuous training and teaching innovation (Perales, Jornet & González, 2014). The design of a new system of global evaluation of teaching quality was then necessary, a system that addresses this functional complexity through diverse tools and multiple sources of

information, thus reflecting the dimension of the teacher's work in this field (Benito & Cruz, 2006; Zabalza, 2003; Zabalza, 2009), a process in which teachers should be involved (Mayor, 2009; Valcárcel, 2003).

In the context of the Canary Islands, its two public universities joined the Docentia Programme. In the case of the University of La Laguna (ULL), the elaboration of the model has added value to the Programme, since the document represents an effort of integration and collaboration among all the agents involved in the adoption of a common framework to the evaluation of teaching activity for the Spanish University System.

The implementation of the Docentia evaluation model thus provided an opportunity to make visible the work of the teaching staff, as well as the new demands and requirements that they would have to assume in the new university context. However, although the final goal was the search for teaching excellence, in compliance with requirements of the model (ANECA, 2007), the document, among other actions linked to the results of the evaluation, makes the following explicit:

a) the design of training plans adapted to the needs of the teaching staff;

b) the assessment of teaching activity and its certification as an compulsory merit in the national accreditation system for access to university teaching staff categories (RD. 1312/2007, of 5 October), and in the competition for access to university teaching staff categories (RD. 1313/2007, of 5 October)

c) its consideration as a merit for the granting of aid for teaching innovation programs and the dissemination of its results, as well as for the scale of merits for recruitment and promotion of the teaching staff, and for the distribution of the budget of Centres and Departments and, finally, the identification of good teaching practices and innovations used by teachers and their dissemination to the rest of the university community (University of La Laguna, 2009).

In essence, it will have to be a useful, precise evaluation, with consequences (Alfageme & Caballero, 2010; Tejedor & García-Valcárcel, 2010) but at the same time a way of stimulating the adoption of good practices that contribute to the improvement of teaching quality (Pozo, Giménez & Bretones, 2009). In short, it is necessary to improve the evaluation of teaching activity in order to become a useful tool for teachers, allowing them to assess the improvement of their teaching methodology, and perceive its strengths and weaknesses (Calderón & Escalera, 2008).

The Docentia model for teacher evaluation: Universidad de La Laguna

The Docentia-ULL programme is a model for evaluating the teaching activity of its teaching staff, which takes into account the activities carried out by teachers before, during and after the teaching of their subject, the results obtained, as well as the activities of training and innovation that he/she has carried out, aimed at teaching improvement. The necessary procedures are adopted to include in the final result of the evaluation the assessment of all the agents involved in the teaching-learning process. In addition, it takes into account the assessment of all those tasks carried out by the teacher in this ambit of his/her teaching performance, which is not the only one, taking into account the complex and diverse set of functions performed by a university teacher.

To this end, this model is based on the assessment of three dimensions: Planning of Education, Teaching Development and Results, which in turn are composed of related sub-dimensions, characterised by a series of indicators with a certain specific weight in the final result of the evaluation, also taking into account the criteria of Adequacy, Satisfaction, Efficiency and Orientation to improvement that characterise the quality of teaching, as stated by ANECA.

The conception of the teacher that underlies this model is that of an agent of change through a range of diverse activities aimed, mainly, at the acquisition of student skills, from the design of the tasks necessary for it, its development and application, to the subsequent analysis and assessment of results, which is the source of feedback, necessary for continuous improvement. In short, teaching evaluation is the tool that allows a final reflection within the Plan, Execute, Revise and Act scheme that proposes the circle of quality and continuous improvement of Deming.

It is essential that the evaluation model be simultaneously effective and accepted by the teacher being assessed, understanding that it constitutes an important tool to reflect on his or her teaching performance (Mayor, 2009; Valcárcel, 2003). To this end, the model, like any other performance evaluation system, must be valid, reliable and have the capacity to discriminate between teachers with favourable results and those with less favourable ones.

To this end, the evaluation of teaching activity is carried out within the framework of the Docentia-ULL programme based on the triangulation of three sources of information: the teacher, students, and academic managers. Specifically, three information collection instruments were designed around indicators of the teaching activity included in the main dimensions of the model, thus allowing a complementary and global evaluation of this part of the academic activity: Teacher Self-Report; Academic managers' Report, and Student Satisfaction Survey (University of La Laguna, 2009).

The evaluation model of the Docentia-ULL programme has to contribute both, the necessary diagnosis of the teaching activity of the teaching staff for their continuous improvement, and the enhancement of the same, facilitating their recognition and professional development. One of the requirements for this is its necessary capacity to discriminate between the four possible results that the participant can obtain: "Unfavourable", "Favourable", "Very Favourable" or "Excellent". However, in line with Isla-Díaz et al. (2014), a common characteristic of most of the assessment models of the teaching quality of Spanish universities through the ANECA Docentia Programme that has been detected is its limited capacity to adequately distinguish "Excellent" teaching staff from the other categories. Table 1 shows the data from 16 Spanish universities whose results are publicly available to this regard.

Table 1. *Comparative results in percentages of teaching staff evaluation in 16 Spanish universities, according to evaluation categories in different calls.*

		Universities																
Academic year	**Qualification**	**Autónoma de Madrid***	**Burgos***	**Girona***	**Politécnica de Valencia***	**Europea Miguel de Cervantes***	**A Coruña**	**Alcalá**	**Cádiz**	**Complutense de Madrid**	**Córdoba**	**Extremadura**	**La Laguna**	**La Rioja**	**Las Palmas de Gran Canaria**	**Salamanca**	**Santiago de Compostela**
2010/11	**E**				18			0		5			90	42		52	79
	MF			100	74			66		10			-	0	100	45	14
	F				6			34		78			9	57		-	0
	D			0	2			0		7			1	1	0	2	5
2011/12	**E**		19		18		30	0	33	1	56	33	94	25		36	
	MF		63	98	74		53	60	67	43	-	36	6	32	100	57	
	F		15		6		13	35		54	44	11	0	44		7	
	D		4	2	2		5	1	0	2	-	20	0	0	0	0	
2012/13	**E**		4			16		0	67	6	62	50	90	42		16	66
	MF		67	100		64		64	33	50	-	33	7	31	99	37	30
	F		29			16		29		43	38	5	1	27		46	2
	D		0	0		4		6	0	2	-	12	2	1	1	0	2
2013/14	**E**	39	31			21	21	9	65		36	41	86	63		20	
	MF	53	55	97		64	68	56	35		-	37	11	29	100	41	
	F	7	10			14	9	31			64	18	1	8		38	
	D	1	4	3		0	3	4	0		-	4	3	1	0	0	
2014/15	**E**		25			17	22	11	76			53	79	56		25	
	MF		73			39	59	61	24			32	18	32	100	42	
	F		2			44	10	26				11	1	12		33	
	D		0			0	9	2	0			4	2	0	0	0	
2015/16	**E**	33	23			43	22	8	57	20			27	60		33	
	MF	58	70			43	59	68	43	78			47	24	100	37	
	F	5	6			14	10	21		2			22	15		20	
	D	2	0			0	9	1	0	1			4	1	0	0	
2016/17	**E**	41	6			27						59	34				
	MF	49	86			46						11	48		100		
	F	7	8			19						26	15				
	D	2	0			8						0	3		0		

E=Excelent, VF=Very favourable, F=Favourable, U=Unfavourable

*: With the model certified by ANECA

Note: In some universities/years the sum of percentages is not exactly 100% due to either rounding or being to the fact of not being included the teachers excluded from the evaluation in the percentages published.

In fact, ANECA, while monitoring the verified models in the different universities, has been insisting on the limited capacity of discrimination of the models, and on the need to reach a greater discriminative capacity in order to be effective. This recommendation is usually presented as necessary to certify the implementation of the model. The requirement of Docentia being a tool that adequately discriminates among teachers, in particular the excellent teachers, seriously affects its prestige. We are referring to the fact that if this evaluation tool leads, for example, to 90% of the teaching staff being qualified as "Excellent", given that theoretically a distribution of the qualification of the teaching staff more centred towards the midpoint is expected, the teaching staff that really stands out from the rest will not be reflected. This "ceiling" effect leads to the discredit of the model, with negative consequences for the

necessary incorporation of the evaluation of teaching by university faculty.

Aims of the study

In this context, and after evaluating the results of the implementation of the Docentia model of the University of La Laguna (Docentia-ULL) in 42.5% of the teaching staff (725) who submitted their teaching activity to evaluation in the first three calls, between 2010 and 2013, the Teaching Quality Evaluation Committee of the University of La Laguna set the objective of finding an evaluation model that was not only reliable, but also valid, that guaranteed the difference between the "Excellent" teaching staff, those teaching staff whom does its work in a notable way, those who do it normally and those who do not reach the minimal standard, thus trying to answer the question: Is Docentia a useful tool? (Isla-Díaz et al., 2014). The Teaching Evaluation Committee of the ULL established that the appropriate strategy for achieving this objective was to reduce the weight of indicators associated with teaching obligations, in favour of Results indicators: students' satisfaction surveys and academic managers' reports, together with activities for teaching training and innovation.

Within the European Higher Education Area, university education, insofar as it is oriented towards the learning of competences by students, entails a process of pedagogical renewal. That requires the participation of the teaching staff in this process and, therefore, an adequate incentive for it, which implies a change in the teaching culture both in the teaching staff and in the Institution (Martínez & Esteban, 2005; Pozo, Bretones, Martos, & Alonso, 2011). The fact that Docentia is a model oriented towards continuous improvement, closely associated with the incorporation of a new approach to university teaching, centred on student learning, necessarily leads to a revision of the model in this sense. The obligations in work performance (e.g., compliance with teaching hours) should not be rewarded, as opposed to the necessary encouragement of relevant aspects of teacher performance, such as educational training and innovation. These are the aspects of teaching which are marking the current commitment to educational quality in universities, of whose teachers seem to be increasingly aware and participative (Caballero & Bolívar, 2015; Perales et al., 2014; Zabalza, 2009). Changes can only be taken to the classroom if they have the support of the group responsible for making them a reality (Valcárcel, 2003).

In relation to our commitment to the Results, the case of the consideration of the program of the subject and the clarity of the evaluation criteria can be illustrative. As previously mentioned, Planning of Education and Teaching Development refers to relevant skills for the performance of teachers, such as the establishment of an appropriate evaluation system (Sinahuya & Sánchez-Tarazaga, 2018) or the adequacy of the taught content with the credits assigned to the subject. Together with the ability to motivate students and adequate tutoring, these skills are considered key to the profile of a good university teacher (Caballero & Bolívar, 2015; San Martín, Santamaría, Hoyuelos, Ibáñez, & Jerónimo, 2014; Tejedor & García-Varcárcel, 2007; Tejedor & García-Varcárcel, 2010; Zabalza, 2009). At the beginning of the implementation of the EHEA, it made a lot of sense to encourage the elaboration of the program and to make the evaluation criteria explicit in the evaluation of teacher performance, given that it requires teachers to put teaching planning into practice. However, once this practice has been consolidated and assumed as a teaching obligation, its incentive loses meaning, inasmuch we are not sure that elaboration of the program implies by itself the capacity of the teaching staff for an adequate planning of teaching. In contrast, these capacities can be considered more reliably valued in the Results section, particularly in the student satisfaction survey (see Pozo Muñoz et al., 2011). More generally, we considered that it made sense to review the verified Docentia ULL Model to identify the 'burned' indicators that had already met their objective, and where little variability was seen along with a 'ceiling' effect.

In relation to the procedure, first we examined the weights of the dimensions of the verified Model, and of the sub-dimensions within each dimension. As mentioned, the initial idea was to give less weight to the "ordinary obligations" of teachers in the evaluation compared to the Results Dimension, and also to their participation in educational training and innovation activities. In this regard, we developed two alternative models for calculating scores, and we are going to compare the data obtained with the model verified by ANECA with data that would be obtained by simulating the results of these alternative models. Subsequently, once the selected alternative model is implemented in the University of La Laguna, we are going to longitudinally monitor the distribution of the real data scores obtained with this new model in the first two calls (2015/16 and 2016/17). In both alternative models to the verified model, the weights of the dimensions and sub-dimensions were modified in order to prioritize the commitment of the teaching staff to training and educational innovation.

Another aspect we considered necessary to revise was the one related to the report of the academic managers and the student satisfaction survey. Both act in the verified Model as requisites for the inclusion of teachers in a certain qualification category. The good performance of the teaching staff must necessarily be translated into the satisfaction it generates, mainly in its students. In fact, in the context of the EHEA, the basic tool for evaluating teacher performance is the student satisfaction survey (Pozo Muñoz et al., 2011).

In addition, we were interested in examining the association between the three sources of evaluation: merits of the teaching staff, satisfaction of students and of academic managers, on which the model for the triangulation of information is based. It is assumed that they are non-redundant and relatively independent sources. The information we obtain would have a certain diagnostic value for our knowledge of the Docentia evaluation model.

In summary, the revision of the verified Docentia ULL Model constituted an opportunity to orient teaching activity according to the quality objectives that correspond to a higher education institution, strategically encouraging the performance of teachers in those activities directly linked to quality, in the context of the change in the educational model that Spanish universities must assume.

Method

Sample

We analysed the scores of the third call for evaluation of the teaching activity for teachers of the ULL in the academic year 2012/13, which brings together a sample of 21.5% of the teaching staff (n=367).

Instruments

The evaluation of the teaching performance of the ULL faculty is carried out through the triangulation of three sources of information: student satisfaction questionnaire, the report of their academic managers and the merits and activities of the teaching staff itself, in the three dimensions of the model that are included in a self-report (University of La Laguna, 2009).

As illustrated in Table 2, the three dimensions of the Teaching Model are: 1. *Planning of Education*, 2. *Teaching Development*, and 3. *Results and Innovation*. *Planning of Education* refers, among other aspects, to the degree of updating and delivery of guides of degree subjects, as well as the level of vertical and horizontal coordination within the degrees in which the teacher participates. *Teaching Development* includes the degree of perceived compliance in relation to activities such as publishing marks, class and tutoring schedules, as well as complementary and special activities carried out by the teacher in order to enrich the learning process of his or her students. The Dimension of *Results and Innovation* includes students' and academic managers' satisfaction, as well as the quantitative evidence referring to the rates of performance, efficiency and abandonment of the subjects taught by the

teacher, together with the activities developed by the teaching staff in relation to their own training and the strategies of teaching innovation carried out in teaching their subjects.

Table 2**.** *Scores of original model and simulated models of teaching staff evaluation according to the dimensions and sub-dimensions of Docentia.*

DIMENSIONES / SUB-DIMENSIONES	MAXIMUM VALUE VERIFIED MODEL	MAXIMUM VALUE MODEL 1	MAXIMUM VALUE MODEL 2
1. PLANNING OF EDUCATION	**25**	**15***	**15***
1.1 Participation in teaching management	9	9	9
1.2 Teaching planning	8	6*	3*
1.3 Coordinating activities	8	8	3*
2. TEACHING DEVELOPMENT	**35**	**35**	**35**
2.1 Ordinary Teaching Activities	20	16*	5*
2.2 Complementary Teaching Activities 2.3 Special Teaching Activities	15	30*	30*
3. RESULTS AND INNOVATION	**40**	**50***	**50***
3.1 Quantitative evidences	10	10	10
3.2. Qualitative evidences (Students survey and Academic managers report)	20	25*	25*
3.3 Training and Innovation 3.4 Other merits	10	20*	20*
TOTAL	**100**	100	100

*Changes made with respect to the verified Model

Source: Manual Docentia ULL (2009) and own elaboration

The final result is established in a maximum of 100 points, which is distributed in dimensions with different weights, and sub-dimensions that contribute to each dimension, with different tops. The sum of subdimensions may exceed the maximum score given to each dimension (in column 2 the weights and tops are shown both in the 2012-13 academic year model called "Verified Model" and in the two simulated models).

It can be observed that dimensions 2.2 and 2.3 are valued together with a maximum of 15 (Verified model) vs 30 (Models 1 and 2). Each sub-dimension counts the corresponding merits, but the sum couldn't exceed the established maximums. The same occurs in dimensions 3.3 and 3.4 which are valued together up to a maximum of 10/20/20 points (depending on the model). *Procedure*

The result of the first certification report issued by ANECA for the verified Docentia ULL Model conveyed the need to discriminate adequately between teachers in order to obtain certification. In addition, during the first three calls, the Evaluation of Teaching Quality Committee detected a considerable level of dissatisfaction, both in the participating faculty and in their academic managers. The dissatisfaction was generated mainly due to the scarce discriminatory power, with a high number of excellent grades, as can be observed in Table 1.

The Committee for the Evaluation of Teaching Quality proposes to study a change in the model in order to comply with the requirements of ANECA, on the one hand, and to improve the satisfaction of the evaluated teachers and their academic managers, on the other.

To this end, and in order to compare the results with the verified Docentia-ULL Model, two simulated models were applied to the matrix of scores of the 367 participants in the third Docentia-ULL call (2012-13 academic year). These two simulated models adjusted the weights of the dimensions, as well as the tops of the sub-dimensions of the original model verified by ANECA. In addition, the minimums to be achieved by the teaching staff to obtain a "Favourable", "Very Favourable" or "Excellent" result were adjusted in the so-called "Qualitative

Evidence", i.e., in the student survey and in the academic managers' assessment.

Given that the merits of dimension 1 *Planning of Education* (see Table 2) basically constitute the obligations of a university teacher; especially the sub-dimension *Teaching planning* (1.2), it is proposed to reduce its weight in favour of increasing other voluntary activities.

Thus, in the first simulated model (Model 1), as illustrated in Table 2, the sub-dimension *Teaching planning* (1.2) is reduced to a maximum of 6 points instead of 8. The maximum of this dimension 1 (P*lanning of Education*) is reduced to a maximum of 15. Along the same line, dimension 2 *Teaching Development* retains its weight; however, the sub-dimension of *Ordinary Teaching Activities* (2.1) drops to 16 points instead of 20, while the *Complementary and Special Teaching Activities* (2.2 and 2.3) together raise their maximum value to 30 instead of 15. Finally, the fact of having decreased by 10 points the dimension 1 P*lanning of Education* allowed the magnitude of dimension 3 *Results and Innovation* to be increased, thus reaching a maximum of 50 points, instead of 40 with the verified Model. *The Qualitative Evidence* sub-dimension (3.2) increases its ceiling from 20 to 25 points and *Training and Innovation* (3.3) together with other merits (3.4) goes from 10 to 20 maximum points.

On the other hand, and as can be seen in the weights of Model 1, the sums of the scores of the sub-dimensions exceed the maximum of each dimension. This allows the teacher to achieve the maximum score for each dimension in different ways, following the principle of equifinality. However, the detail of the changes for the sub-dimensions in Model 2, as can be seen in Table 2, is more restrictive.

On the other hand, the threshold for obtaining a "Very favourable" result increases to 70 points, the minimum score required in the student satisfaction survey is 9.8 out of 15 possible points (equivalent to 6.5 out of 10 points) and in the report of academic managers the minimum is also established at 6.5 out of 10 possible points. For an "Excellent" result, between 90 and 100 points must be achieved, with a minimum result in student satisfaction of 12 out of 15 possible points (equivalent to 8 out of 10) and in the report of the academic managers also of 8 out of 10 possible points (see Table 3). Likewise, the conditions for obtaining a "Favourable" result would be, in addition to obtaining 50 points in the evaluation, to achieve a minimum score of 7.5 out of 15 (equivalent to 5 out of 10) in the student satisfaction survey, as well as in the assessment report of academic managers.

Table 3. *Qualification of teaching staff as a function of the score obtained both in the verified and in simulated Models*

Teacher qualification	Verified model	Simulated models
Excellent	80 -100 points	90-100 points Students' Satisfaction: 12-15 (~ 8/10 – 10/10) Academic manager's Report:8-10
Very Favourable	65 – 79,9 points	70 – 89 points Students' Satisfaction: 9.8-11.9 (~ 6.5/10 – 7.9/10) Academic manager's Report: 6.5-7.9
Favourable	50 - 64,9 points	50-69 points Students' Satisfaction: 7.5 – 9.7 (~ 5/10 – 6.4/10) Academic manager's Report: 5-6.4
Unfavourable	<50 points	<50 points

Source: Manual Docentia ULL (2009) and own elaboration

In the second simulated model (Model 2), most demanding, in addition to intervals and minimum adjustments applied in Model 1, the sub-dimensions *Teaching Planning* (1.2) and *Coordinating Activities* (1.3) are adjusted to have a maximum of 3 points and the sub-

dimensions of *Ordinary Teaching Activities* (2.1) are adjusted to 5 (see Table 2).

Both Models made more flexible the ways in which a teacher could reach the maximum score in each dimension, while at the same time preventing him/her from accumulating all the merits in one dimension, to the detriment of the others. At the same time, it was made harder to achieve the rating of Favourable, as well as the qualitative leaps towards Very Favourable or Excellent, through the minimum score of its students and the Academic managers.

After analysing the results of the simulations, the Committee for the Evaluation of Teaching Quality agreed upon with the labour representatives of the Teaching and Research Staff to select Model 1 to be applied in the second five-year phase of the evaluation of the teaching staff. This consensus was based on the excessive demand of Model 2, whilst Model 1 already met the stated objectives.

Results

Simulated models and discriminability in teacher evaluation

Figure 1 shows the distribution of the teaching staff by rating category according to the verified Model and each of the simulated models.

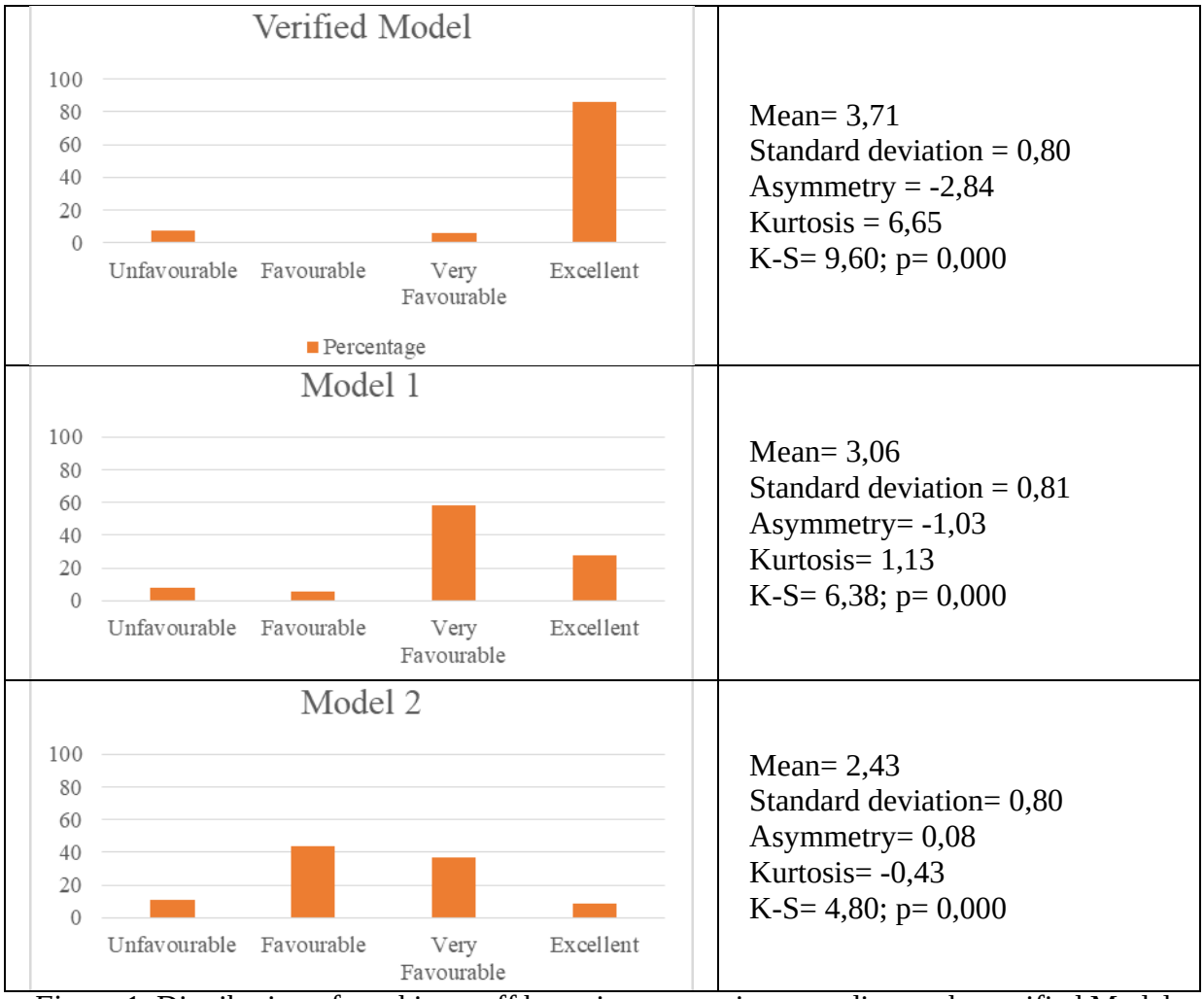


Figure 1. Distribution of teaching staff by rating categories according to the verified Model and the two simulated models

As can be seen, the modifications carried out in the simulated models produce a notable decrease in the number of teachers rated as "Excellent" in the so-called Model 1, with a somewhat more centred distribution (asymmetry index = -1.03). Nevertheless, like the verified Model (asymmetry index = -2.84), it maintains a bias towards the positive

evaluation of the faculty, in this case focused on the "Very favourable" category. In contrast, Model 2, which is more demanding when it comes to giving greater weight to the *Results and Innovation* dimension, together with greater restrictions on the weight of the *Planning of Education* dimension, as well as on the sum of the ceilings of the different sub-dimensions that contribute to it, leads to a more centred distribution (asymmetry index = 0.08) of the teaching staff in the different evaluative categories. In short, the contrast between the Model verified by ANECA for the ULL, and the simulations carried out, indicate that Docentia-ULL is capable of discriminating the teaching staff, adhering to a distribution of the scores with less asymmetry to the right with Model 1, or to a distribution centred on the intermediate categories (Model 2). None of the three models is distributed normally, although it can be seen that Model 1 and Model 2 are closer to the normal curve than the verified Model.

Although the results of the simulation of Model 2 are closer to normality, the ULL opted for the implementation of Model 1, less demanding than Model 2, but with the discriminatory capacity between the ratings of the teaching quality demanded by ANECA and agreed, as explained above.

Figure 2 compares the distribution of teacher qualification in each of the simulated models with the real data of the verified Model and the real data obtained in the 2015/16 and 2016/17 calls by applying the alternative model (Model 1). We can see how the real results obtained with the implementation of Model 1 are close to the distribution of the simulated Model 1, fundamentally as regards the categories of "Excellent" and "Very favourable". In addition, they show the ability to discriminate more adequately between "Favourable", "Very Favourable" and "Excellent" faculty.

In this sense, the Chi-square contrast of goodness-of-fit is significant if we compare the distribution of Model 1 with the distributions obtained in the 2015/16 and 2016/17 academic years (χ=50.55; p=0.00 and χ=21.38; p=0.00, respectively). In both editions we get more "Favourable" faculty than expected. However, if we make this contrast without taking into account the "Favourable" category, the distributions obtained for the 2015/16 and 2016/17 courses no longer differ from the expected distribution according to Model 1 (χ=1.75; p=0.42 and χ=5.95; p=0.051, respectively).

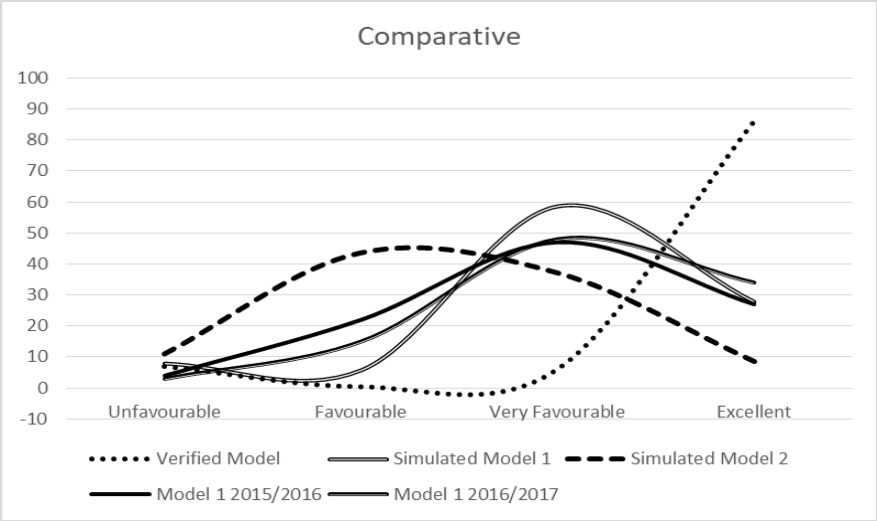
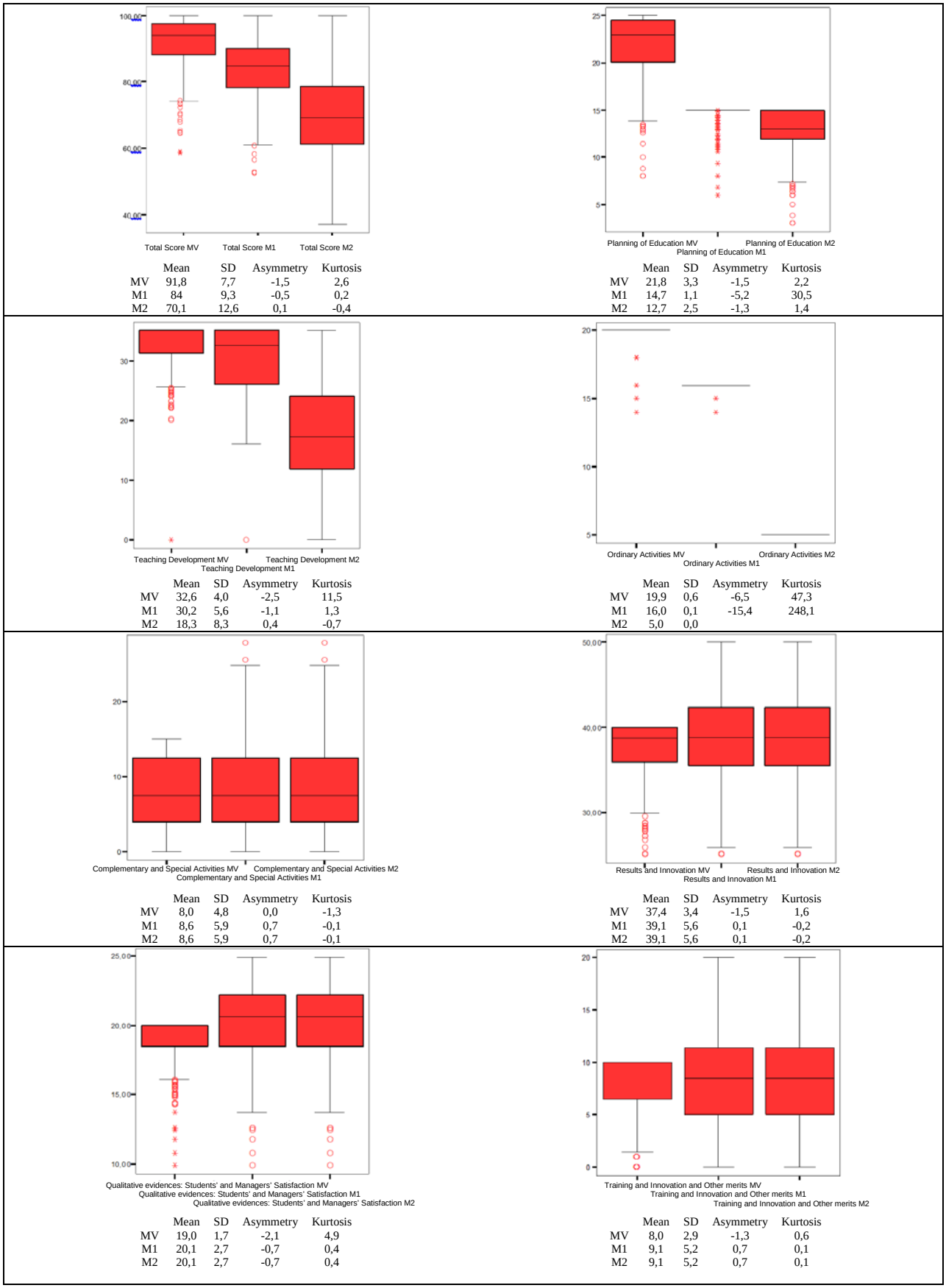


Figure 2**.** Comparison of the Distribution of teaching staff by category (%) according to the verified Model, the simulated and the real results obtained after applying Model 1 in 2015/16 and 2016/17.

Figure 3 shows the distribution and synthesis indicators of the different dimensions and sub-dimensions according to the Verified Model, simulated Model 1 and Model 2 obtained by resulting from the change in criteria presented in the procedure.



	Mean	SD	Asymmetry	Kurtosis
MV	91,8	7,7	-1,5	2,6
M1	84	9,3	-0,5	0,2
M2	70,1	12,6	0,1	-0,4

	Mean	SD	Asymmetry	Kurtosis
MV	21,8	3,3	-1,5	2,2
M1	14,7	1,1	-5,2	30,5
M2	12,7	2,5	-1,3	1,4

	Mean	SD	Asymmetry	Kurtosis
MV	32,6	4,0	-2,5	11,5
M1	30,2	5,6	-1,1	1,3
M2	18,3	8,3	0,4	-0,7

	Mean	SD	Asymmetry	Kurtosis
MV	19,9	0,6	-6,5	47,3
M1	16,0	0,1	-15,4	248,1
M2	5,0	0,0		

	Mean	SD	Asymmetry	Kurtosis
MV	8,0	4,8	0,0	-1,3
M1	8,6	5,9	0,7	-0,1
M2	8,6	5,9	0,7	-0,1

	Mean	SD	Asymmetry	Kurtosis
MV	37,4	3,4	-1,5	1,6
M1	39,1	5,6	0,1	-0,2
M2	39,1	5,6	0,1	-0,2

	Mean	SD	Asymmetry	Kurtosis
MV	19,0	1,7	-2,1	4,9
M1	20,1	2,7	-0,7	0,4
M2	20,1	2,7	-0,7	0,4

	Mean	SD	Asymmetry	Kurtosis
MV	8,0	2,9	-1,3	0,6
M1	9,1	5,2	0,7	0,1
M2	9,1	5,2	0,7	0,1

Figure 3. Distribution and synthesis indicators of the different dimensions and sub-dimensions according to the Verified Model, simulated Model 1 and Model 2.

It can be observed that the distribution of the total score increases its variability, decreasing its mean, asymmetry and kurtosis. The first dimension, *Planning of Education* (compulsory activities), receives a 15-point ceiling effect. Dimension 2, *Teaching Development*, also increases its variability, decreasing its mean, asymmetry and kurtosis. It is composed of the sub-dimensions *Ordinary Teaching Activities*, which receive a ceiling of 16 or 5 points depending on which model and *Complementary Teaching Activities and Special Activities*, which manage to increase their mean and variability. Dimension 2, *Results and Innovation*, increases its range of scores, slightly increasing its mean and increasing its variability. Its sub-dimension Qualitative evidence increases its range, in the same line as the sub-dimensions *Training and Innovation* and other merits, which increases its variability and its mean by increasing its overall ceiling.

Regression and correlational data on the relationship between the three sources of evaluation: merits of the teaching staff, satisfaction of students and academic manager's satisfaction.

The Docentia Model assumes that the three sources of information for teacher evaluation are independent, that is, they are not redundant. To examine this assumption, we have correlated *Student's Satisfaction*, the *Satisfaction of Academic managers* and several dimensions/sub-dimensions referring to the merits of the teacher: *Planning of Education, Teaching Development* and the sub-dimension of *Training and Innovation*, both in the verified Model and in Model 1 (which was implemented). The correlations are shown in Table 4.

Table 4. *Correlations between dimensions/sub-dimensions of teacher's merits with student's and academic managers' satisfaction, both in the verified Model (MV) and Model 1 (M1).*

	Ordinary teaching Activities	Complementary and Special Activities	Training Innovation and Other merits	Students' Satisfaction	Academic managers	Total Score
Planning of Education MV	0.080	0.334	0.367	-0.089	0.114	0.699
Planning of Education M1	-0.018	0.209	0.200	-0.030	0.044	0.392
Ordinary teaching Activities MV		0.005	0.055	0.084	0.039	0.087
Ordinary teaching Activities M1		-0.031	0.015	0.016	-0.005	-0.028
Complementary and Special teaching Activities MV			0.274	0.066	0.247	0.585
Complementary and Special teaching Activities M1			0.277	0.051	0.226	0.639
Training and Innovation and Other merits MV				-0.054	0.134	0.628
Training and Innovation and Other merits M1				-0.067	0.088	0.715
Students' Satisfaction MV					0.068	0.170
Students' Satisfaction M1					0.068	0.201
Academic Mangers MV						0.303
Academic Mangers M1						0.306

n=367; correlaciones significativas al 0,05 a partir de un valor de 0,114.

As can be observed in this Table, *Student's Satisfaction* is independent of the merits of the teacher and of the assessment of the academic managers, both in the verified Model and in Model 1. As for the *Satisfaction of the Academic managers*, it has a low correlation both with the complementary teaching activities in both models, and with educational training and innovation. These data support the independence of the three sources on which the model for triangulation of information is based: merits of the teacher (self-report), satisfaction of the students and the evaluation of the academic managers.

However, although a certain independence of the three sources is positive for the validation of the Docentia Model, it is certain that a low to moderate relationship between the three

sources could be expected; but this only occurs in the case of the correlation between *Complementary Teaching Activities* and the *Academic managers' Satisfaction*. It is particularly striking to us that *Training and Innovation* does not have a higher correlation with *Student Satisfaction*.

In order to know the weight of the merits of the teacher, as well as the S*atisfaction of the Students* and the *Academic managers* in the total score obtained, we have carried out linear regression analysis, both for the verified Model and for Model 1 (Table 5).

Table 5. *Regression coefficients, significance and semipartial correlation of the sub-dimensions on the total score obtained in the verified Model (MV) and in Model 1 (M1).*

	B	Standard error	Beta	t	Sig.	Semi-partial
Planning of Education MV	1.090	.061	.473	17.877	.000	.421
Planning of Education M1	1.616	.189	.196	8.535	.000	.189
Ordinary teaching Activities MV	.060	.305	.005	.198	.843	.005
Ordinary teaching Activities M1	-1.876	1.777	-.023	-1.056	.292	-.023
Complementary and Special teaching Activities MV	.449	.042	.282	10.811	.000	.255
Complementary and Special teaching Activities M1	.625	.038	.397	16.576	.000	.367
Training and Innovation and Other merits MV	.976	.068	.372	14.423	.000	.340
Training and Innovation and Other merits M1	1.014	.042	.568	24.276	.000	.537
Students' Satisfaction MV	.707	.082	.205	8.575	.000	.202
Students' Satisfaction M1	.897	.093	.215	9.655	.000	.214
Academic Mangers MV	.650	.137	.116	4.736	.000	.112
Academic Mangers M1	.968	.155	.142	6.242	.000	.138

MV: R^2=0.80; $F_{6,360}$=240.078; p= 0.000.
M1: R^2=0.82; $F_{6,360}$=280.312; p= 0.000.

As can be seen in Table 5, for both models the determination coefficient (R^2) is significant, explaining 80 and 82% of the variability of the total score for the verified Model and Model 1, respectively. *Ordinary teaching activities* (imperative) are the only predictor that is not significant in any model. The contribution (ß) of *Complementary Teaching Activities* and that of *Training and Innovation* is greater in Model 1 than in the verified Model. On the other hand, the merits of the teaching staff associated with *Teaching planning* lose their weight in Model 1, in favour of *Training and Innovation*. Along the same lines, semi-partial regression coefficients could be interpreted.

These results support our starting assumption: teaching obligations are not useful to discriminate between favourable and excellent teachers. Likewise, it supports that in Model 1, the "voluntary" merits of the teaching staff acquire a greater weight with respect to the verified Model and that the obligations lose their weight, which was one of the objectives we were pursuing.

Discussion

The results obtained with the implementation of the alternative model (Model 1) in the first two calls of the new five-year Docentia ULL (2015/16 and 2016/17) show its discriminatory capacity once adjusted to the proposed modifications based on the simulations carried out.

From the study of the distributions of the models, it can be deduced that the weight given in the verified Model to the dimensions and sub-dimensions linked to the teaching obligations (sub-dimension 1.2: *Teaching planning*; sub-dimension 2.1: *Ordinary Teaching Activities*), as well as the flexibility in the tops of these sub-dimensions to reach the maximum score of each dimension (1. *Planning of Education* and 2. *Teaching Development*) explains to a large extent the bias towards the most positive categories. In contrast, the increase in weights, and also the

restrictions on obtaining the maximum score in dimension 3. *Results and Innovation*, leads to a clear decrease in the number of teachers qualified as "Excellent".

When less weight is assigned to teaching obligations and more weight is given to voluntary merits, that is, their continuous training and their involvement in teaching innovation, a more focused distribution in the intermediate categories of teachers is obtained, discriminating positively "Excellent" teachers. Continuous training and innovation in the teaching work are marking the current commitment to the educational quality of universities. However, as we have already commented, changes in the teaching-learning model can only be brought into the classroom if they have the support of the group responsible for making them a reality (Valcárcel, 2003). Awareness of their value is being incorporated among teachers, which is associated with their greater participation in teaching projects of this nature (see Perales et al., 2014). Even so, the effort must continue, given that the experiences are dispersed, and the universities do not have systematic mechanisms that make possible the extension of good educational practices.

On the other hand, what is theoretically expected, according to the performance model that is proposed in Docentia for a collective such as university faculty, is a distribution that is not as centred as that caused by Model 2, but less asymmetric than the verified Model. This expected distribution is the one found in Model 1 of our study. Therefore, Docentia-ULL is able to discriminate between teachers who fulfil their teaching obligations in a "Favourable" way and those who do so in a "Very Favourable" or "Excellent" way, once the model is adopted based on the results found with the simulations. With this, the model can recover its prestige among teachers and its capacity to orient the activity of teachers towards the commitment to excellence. The change of educational model from the teacher-centred master class to the figure of the teacher as a dynamic agent in a student-centred learning process requires this commitment.

As previously mentioned, the model is based on the triangulation of three sources of information: student satisfaction, satisfaction of academic managers and merits of the teaching staff. It is assumed that they are non-redundant and relatively independent sources, which has been supported by the correlational analysis we have carried out. It could be argued, in the opposite sense, that satisfaction, especially of the students, should present a moderate association with the effort of the teaching staff in areas such as *Training/innovation* or *Teaching Development*. Greater partnership might even be desirable. According to previous literature, students positively value teacher skills such as the establishment of an appropriate evaluation system, the appropriateness of the content taught with subject credits, the ability to motivate or adequate tutoring. These competencies are considered central to the profile of university teachers (Caballero & Bolívar, 2015; Pozo Muñoz et al., 2011; San Martín et al., 2014; Tejedor & García-Varcárcel, 2007; Tejedor & García-Varcárcel, 2010; Zabalza, 2009). The innovation capacity of the teaching staff would be reflected in the satisfaction of the students meanwhile it results in motivation and stimulus for study and facilitate their learning. In this respect, a positive association of educational innovation practices with student satisfaction has been observed, insofar as they positively affect motivation and learning, when innovation refers to the use of ICT for teacher improvement (García-Varcárcel & Tejedor, 2017; Herrero, 2014).

The absence of a relationship between such a relevant factor in teacher performance, as teacher training and their innovative capacity, with student satisfaction deserves reflection in our study. Firstly, with regard to its implications for the reform of the Docentia model that we have proposed, which gives greater weight to educational training and innovation in teacher evaluation. If it turns

out that it does not result in student satisfaction, the greater weight we have given it could cause a lack of convergent validity. On the other hand, the low correlation found may be due to a methodological problem of our study. The student satisfaction survey of the ULL (as in the majority of Spanish universities) does not contain items that directly assess whether the teaching staff carries out educational innovations in the teaching they give or satisfaction with them. Therefore, satisfaction with training and educational innovation can only be measured, assuming that it is related to other capacities of the teaching staff that are measured (such as the capacity to motivate students or the use of didactic resources that facilitate learning). In this regard, a relevant question that we can ask ourselves is whether specific items on the innovation capacity of teachers and satisfaction with it should be included in the student survey. This would allow a direct assessment of whether training and innovation are reflected in student satisfaction.

By way of speculation, the following are some of the reasons that may be influencing this low association. In this respect, it may happen that educational innovation does not always affect the motivation or perception of students as facilitators of their learning (and academic success) and, therefore, their satisfaction with teaching. This could be the case of educational innovation that addresses more sophisticated assessment strategies (such as the use of rubrics and assessment guides) which, although relevant to the assessment of competence acquisition, may not have a direct motivating effect on students (they are relatively unrelated to their interests), and where their relationship with improving learning is not so evident. At the same time, they may entail a greater effort on their part or be redundant with traditional evaluation, which are negative elements for the assessment of their satisfaction with the teachers who implement them. In fact, in a previous research carried out in the context of the Andalusian university system, it was found that an item in the student body survey that can be linked to teacher training and educational innovation "Uses didactic resources that facilitate learning" was not a significant predictor of overall student satisfaction (Pozo, Giménez, & Bretones, 2009).

In the current context, educational innovation in our university is based on particular teacher initiatives, is abundant, but often unequal, and affects diverse areas such as the use of ICT, teaching methodology or evaluation. We consider that their association with student satisfaction is also uneven, which could help explain our correlational results. As far as we know, there are no previous studies that systematically analyse the relationship between educational innovation, in all its amplitude, and student satisfaction. Therefore, carrying out new research that investigates the relationship between educational innovation and teacher training with student satisfaction seems appropriate.

Teaching activity evaluation systems need to be improved so that evaluation becomes a useful instrument for teachers and guides them towards improving their performance. In the ULL this study has constituted a starting point for the modification of the verified Model, after its implementation during the first five years. As a result of this experience, Docentia-ULL has been modified in the sense that it has indicated to us. In agreement with the labour representatives of the teaching staff, and with the support of the institution, Model 1 has been implemented since the call for Docentia-ULL 2015/16, which has shown to have the discriminatory capacity sought. The objective has institutional relevance, since a better and more adequate discrimination of the teaching staff, according to the bases of the Docentia Model, gives credibility and prestige to the teaching work. In this respect, it is necessary that institutional support continues so that the evaluation of teacher performance has institutional and personal effects in the promotion of the teaching staff. So, it is convenient to bear in mind the recommendations of the "High

Group on Modernization of Higher Education" of the European Commission of 2013, in the sense of balancing the institutional relevance that is given to teaching with that of research, to all effects (Caballero & Bolívar, 2015; Zabalza, 2009). A decisive impulse to educational training and innovation becomes really necessary and, with it, to carry out policies aimed at the incorporation of the culture of educational innovation among our students and teachers (Roig-Vila, 2017). It could also be relevant to bet on models of teaching-learning and evaluation that have a certain unifying character; particularly the real support to continuous evaluation would allow a greater unification of educational innovation practices (Fraile, López-Pastor, Castejón, and Romero, 2013; Hortigüela, Pérez-Pueyo, and López-Pastor, 2015). Likewise, it would be opportune for the institution to carry out policies aimed at giving prestige to educational innovation, and to the teachers who carry them out, so that the capacity of the teacher can be adequately transferred to the satisfaction of the students.

This study has the limitation of being focussed to a single university. However, we consider that it has potential for extrapolation to other universities, given that the Docentia Model, in particular its evaluative dimensions and the criteria of Adequacy, Satisfaction, Efficiency and Orientation to Improvement that determine its contents, constitute a common framework in the Spanish University System (SUE). As can be seen from Table 1, the difficulty in distinguishing "Excellent" teaching staff is common to a great number of Spanish universities. In any case, it is necessary to carry out studies in other SUE universities in order to examine the effectiveness of the changes in our simulated models for an adequate teaching staff categorization. As well, it is important to continue monitoring the results obtained in the ULL with the new Docentia-ULL model during the five-year period that has already begun, for a new revision.

On the other hand, it is necessary to generalise this kind of reflections throughout the SUE, and that the quality agencies (ANECA and autonomous agencies) take them on, whereas the evaluation of the teaching staff affects their own promotion. Teacher's accreditation procedures, such as ACADEMIA, consider Docentia to be a valid instrument for this purpose. If excellence is obtained more easily in some universities than in others, the promotion of some will be given priority over others, without this discrimination being considered either fair or valid, but rather as an arbitrary evaluation of the teaching performance of university teaching staff.

References

Alfageme, M. B., & Caballero, K. (2010). Evaluación y profesorado en la universidad española. *Educatio Siglo XXI, 28*(1), 271-294.

Álvarez-Rojo, V., Asensio-Muñoz, I., Clares, J., Del-Frago, R., García-Lupión, B., & García-Nieto, N. "…" Salmerón, P. (2009). Perfiles docentes para el espacio europeo de educación superior (EEES) en el ámbito universitario español. *RELIEVE, 15*(1). doi: https://doi.org/10.7203/relieve.15.1.4187

Agencia Nacional de Evaluación de la Calidad y Acreditación (2007). *Programa DOCENTIA: Programa de Apoyo a la Evaluación de la Actividad Docente del Profesorado.* Available from http://www.aneca.es/Programas-de-evaluacion/Evaluacion-institucional/DOCENTIA

Agencia Nacional de Evaluación de la Calidad y Acreditación (2008). *Programa Academia. Principios y orientaciones para la aplicación de los criterios de evaluación.* Available from http://www.aneca.es/Programas-de-evaluacion/Evaluacion-de-profesorado/ACADEMIA

Benito, A. & Cruz, A. (2006). *Nuevas claves para la docencia universitaria en el Espacio*

Europeo de Educación Superior. Madrid: Narcea.

Caballero, K., & Bolívar, A. (2015). El profesorado universitario como docente: hacia una identidad profesional que integre docencia e investigación. *REDU: Revista de Docencia Universitaria*, *13*, 57-77.

Calderón, C., & Escalera, G. (2008). La evaluación de la docencia ante el reto del Espacio Europeo de Educación Superior (EEES). *Educación XX1*, *11*, 237-256. doi: https://doi.org/10.5944/educxx1.11.0.316

High Group on Modernization of Higher Education (2013). *Improving the quality of teaching and learning in Europe's higher education institutions*. Available from http://ec.europa.eu/dgs/education_culture/repository/education/library/reports/modernisation_en.pdf

Fraile, A., López-Pastor, V., Castejón, J., & Romero, R. (2013). La evaluación formativa en docencia universitaria y el rendimiento académico del alumnado [The education evaluation in university teaching and the academic performance of students]. *Aula Abierta*, *41*, 23-34.

Herrero, R. (2014). El papel de las TIC en el aula universitaria para la formación en competencias del alumnado. *Pixel-Bit. Revista de Medios y Educación*, *45*, 173-188.

García-Valcárcel, A., & Tejedor, F. J. (2017). Percepción de los estudiantes sobre el valor de las TIC en sus estrategias de aprendizaje y su relación con el rendimiento. *Educación XX1*, *20*(2), pp. 137-159. doi https://doi.org/10.5944/educxx1.19035

Hortigüela, D., Pérez-Pueyo, A., & López-Pastor, V. (2015). Implicación y regulación del trabajo del alumnado en los sistemas de evaluación formativa en educación superior. *RELIEVE*, *21*(1), pp. 1-5. https://doi.org/10.7203/relieve.21.1.5171

Isla-Díaz, R., Ramos, T., Noda, A., Jiménez, F., Acosta, S., Marrero Hernández, H., Pérez Monteverde, M.V., & Soriano, M.I. (2014). Evaluación de la calidad docente: ¿Es el Docentia una herramienta inútil? *XI Foro Internacional sobre Evaluación de la Calidad de la Investigación y la Educación Superior*. July 8 – 11, 2014. Bilbao.

Martínez, M., & Esteban, F. (2005) Una propuesta de formación ciudadana para el EEES. *Revista Española de Pedagogía*, *63*: *230*, January-April, 63-73.

Mayor, C. (2009). Nuevos retos para una Universidad en proceso de cambio: ¿Pueden ser los profesores principiantes los protagonistas? *Revista de Currículum y Formación del Profesorado*, *13*(1), 61-77. Available from https://recyt.fecyt.es/index.php/profesorado/article/view/41904/23910

Murillo, J. (2008). La evaluación del Profesorado Universitario en España. *Revista Iberoamericana de Evaluación Educativa*, *1*(3), 29-45.

Perales, M. J., Jornet, J. M., & González, J. (2014). Tendencias en las políticas de formación y evaluación del profesorado en la Educación Superior en España. *Revista Iberoamericana de Evaluación Educativa*, *2*, 53-64

Pozo, C., Giménez, M. L., & Bretones, B. (2009). La evaluación de la calidad docente en el nuevo marco del EEES. Un estudio sobre la encuesta de opinión del Programa DOCENTIA-ANDALUCÍA. *XXI Revista de Educación*, *11*, 43-64.

Pozo, C; Bretones, B.; Martos, M. J., & Alonso, E. (2011). Evaluación de la actividad docente en el Espacio Europeo de Educación Superior: un estudio comparativo de indicadores de calidad en universidades europeas. *Revista Española de Pedagogía*, *248*, 145-163.

Roig-Vila, R. (Ed.) (2017). *Investigación en docencia universitaria: Diseñando el futuro a partir de la innovación educativa*. Barcelona: Editorial Octaedro.

San Martín, S., Santamaría, M., Hoyuelos, F. J., Ibáñez, J., & Jerónimo, E. (2014). Variables definitorias del perfil del profesor/a universitario/a ideal desde la

perspectiva de los estudiantes pre-universitarios/as. *Educación XX1*, *17*(2), 193—215. doi: https://doi.org/10.5944/educxx1.17.2.11486

Sinahuya, A., & Sánchez-Tarazaga, V. (2018). La competencia evaluativa de los docentes: formación, dominio y puesta en práctica en el aula. *Revista Iberoamericana de Educación*, *76*(2), 95-116

Tejedor, F. J., & García-Valcárcel, A. (2007). Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos). Propuestas de mejora en el marco del EEES. *Revista de Educación*, *342*, 443-473.

Tejedor, F. J., & García-Valcárcel, A. (2010). Evaluación del desempeño docente. *Revista Española de Pedagogía*, *247*, 439-460.

Universidad de La Laguna (2009). *Manual para la Evaluación de la Actividad Docente (DOCENTIA-ULL).* Available from https://web.archive.org/web/20120413024947/https://www.ull.es/Private/folder/institucional/ull/calidad/DOCENTIA/Manual.pdf

Valcárcel, M. (2003). *La preparación del profesorado universitario español para la Convergencia Europea en Educación Superior*. EA2003-0040 Project. Available from http://campus.usal.es/~ofeees/ESTUDIOS_INFORMES_GRALES/informe_final.pdf

Zabalza, M. A. (2003). *Competencias docentes del profesorado universitario. Calidad y desarrollo*. Madrid: Narcea.

Zabalza (2009). Ser profesor universitario hoy. *La cuestión universitaria*, *5*, 69-81.

Authors / Autores	To know more / Saber más
Isla-Díaz, Rosa (risladia@ull.es). Tenured professor of Social Psychology in the Department of Cognitive, Social, and Organizational Psychology at the University of La Laguna. Postal Address: Department of Psychology and Speech Therapy. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain).	ORCID 0000-0003-0252-1873 Google académico
Marrero-Hernández, Hipólito (hmarrero@ull.es). Chair of Basic Psychology in the Department of Cognitive, Social, and Organizational Psychology at the University of La Laguna. Postal Address: Department of Psychology and Speech Therapy. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain).	ORCID 0000-0002-6008-3587 Google académico ResearchGate
Hess-Medler, Stephany (sthess@ull.edu.es). Tenured professor of Methodology of Behavioural Sciences in the Department of Clinical Psychology, Psychobiology, and Methodology at the University of La Laguna. Postal Address: Department of Psychology and Speech Therapy. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain)	ORCID 0000-0002-0289-8796 Google académico ResearchGate
Soriano, Mabel (msoriano@ull.es). Tenured Professor in the Department of Chemical Engineering and Pharmaceutical Technology at the University of La Laguna. Postal Address: Department of Pharmacy. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain).	

Isla-Díaz, Rosa; Marrero-Hernández, Hipólito; Hess-Medler, Stephany; Soriano, Mabel; Acosta-Rodríguez, Severo; Pérez-Monteverde, María-Victoria, & Blanco-Freijo, Marcos (2018). A longitudinal look: Is "Docentia" an useful instrument for university teacher's evaluation? *RELIEVE*, *24*(2), art. 2. doi http://doi.org/10.7203/relieve.24.2.12142

Acosta-Rodríguez, Severo (sacosta@ull.es).

Tenured professor in the Department of Fine Arts at the University of La Laguna. Postal Address: Department of Fine Arts. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain).

Pérez-Monteverde, María-Victoria (mvperez@ull.edu.es).

Department of Economics, Accounting, and Finance at the University of La Laguna. Postal Code: Department of Economics, Business, and Tourism. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain).

Blanco-Freijo, Marcos (mblanco@ull.edu.es).

Quality Unit at the University of La Laguna. Postal Address: Torre Agustín Arévalo. Box 456. 38200-San Cristóbal de La Laguna, Tenerife (Spain).

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa

E-Journal of Educational Research, Assessment and Evaluation

[ISSN: 1134-4032]