Questionnaire evaluating teaching competencies in the university environment. Evaluation of teaching competencies in the university

Juan Antonio Moreno-Murcia[1*], Yolanda Silveira Torregrosa[2], Noelia Belando Pedreño[3]

[1]Department of Health Psychology, Miguel Hernández de Elche University {j.moreno@umh.es}

[2]Department of Health Psychology, Miguel Hernández de Elche University {silveiratorregrosa.y@gmail.com}

[3]Department of Health Psychology, Miguel Hernández de Elche University {nbelando@umh.es}

ABSTRACT

The objective of this study was to design and validate a measuring instrument to evaluate the performance of university professors. The Evaluation of Teaching Performance (CEID [Centro de Estudios e Investigaciones Docentes (Center for Teaching Studies and Research)]) questionnaire was administered to 1297 university students. Various factor analyses were performed (exploratory and confirmatory), of the internal consistency, descriptive statistics, and correlation of all of the items. The data obtained confirmed a suitable psychometric structure for the CEID scale, which was made up of three dimensions (planning, development, and result). It is therefore concluded that it is a valid and reliable instrument for evaluating the performance of a university professor.

KEYWORDS: TEACHING EVALUATION, MEASURING INSTRUMENT, ESTIMATION, UNIVERSITY TEACHING, TEACHING COMPETENCE.

1 INTRODUCTION

Currently, as a result of the adaptation of studies to the European Higher Education Area (EHEA), university education requires a series of important modifications to what has been done until now. This adaptation requires universities to improve the quality of their activities by means of systematic evaluation mechanisms, both internal and external, with one of the pillars of this adaptation being support for the scientific-technical innovation and development processes, which involves a series of organizational and pedagogical changes that could encourage interest in innovative pedagogical methods (Marqués, 2008). In this process of the methodological renovation of universities, the assessment of the students could be the determining factor in finding out and acting on the educational quality of the professors, for which it would be necessary to design measuring instruments in which the dynamic of the teaching-learning process is evaluated in terms of methodology and teaching competencies.

One of the most significant changes in the new model is the formulation of new qualifications based on professional competencies (Carreras, 2005). This focus on competencies overrides the traditional university preparation system based exclusively on subject matters or disciplines. Preparation based on professional competencies aims to provide learning with significance and functionality, and to put into action a set of knowledge, attitudes, skills, and values in the development of professional activity, involving the student in an active manner, and thereby encouraging reflection and critical thinking, allowing the student to make decisions, solve problems, design projects, etc. That is, the use of self-regulated styles that encourage autonomy would facilitate the integration of the knowledge that the student acquires (Leary, Walker, Shelton, & Harrison, 2013), leading the student to show greater interest in and commitment to the subject matter he or she is studying, and to improve his or her academic performance (Vansteenkiste, Soenens, Sierens, Luyckx, & Lens, 2009). This new model involves a significant change in the way the teaching-learning process is conceived at universities, in which the professor must not only design and present the contents to be developed. It is a matter of providing incentive for teacher involvement and developing the critical knowledge of the students, to which end they must specify the teaching methodologies and evaluation systems to be developed for each subject matter in relation to the previously established professional competencies, thereby facilitating their transition to the professional environment (Salmerón, 2013). According to this preparation model based on competencies, the evaluation systems and teaching methods are structured and designed according to the competencies to be acquired, and so working in the new university model will mean greater demands on the part of the faculty (Pallisera, Fullana, Planas, & del Valle, 2010). Incorporating the dynamics of internal evaluation into the management of qualifications will therefore especially require a high level of motivation in the faculty. In this sense, the evaluation of teaching activity appears to acquire a significant value within the new systems,

---
*To whom correspondence should be addressed:
Universidad Miguel Hernández de Elche, Centro de Investigación del Deporte
Avenida de la Universidad, s/n. - 03202 Elche (Alicante)

guaranteeing the quality of university institutions. And we can therefore understand these evaluation processes as a means of reflection to implement policies to act on and improve the quality of teaching performance. In this mission, in order to achieve an improvement in quality in the manner affirmed by the European Association for Quality Assurance in Higher Education (ENQA), university institutions must develop and make available formal mechanisms such as approval, periodic review, and monitoring of their curricula and degrees, using control systems such as objective evaluation.

There is some precedent in Spain for evaluating the quality in university education, such as the *Programa de Evaluación de la Calidad Docente* [Teaching Quality Evaluation Program] (1980), the *Programa Experimental de Evaluation de la Calidad del Sistema Universitario* [Experimental Program for the Evaluation of the Quality of the University System] (1993,1994), and starting in 1996 universities have evaluated the quality of their qualifications through programs such as the *Plan Nacional de Evaluation de la Calidad de las Universidades* [National Plan for the Evaluation of the Quality of Universities], the *II Plan de Calidad de las Universidades* [University Quality Plan II], the *Programa de Evaluation Institucional de ANECA*[i] [ANECA Institutional Evaluation Program] (2007), or by means of other similar independent programs (ACSUG[ii], AVAP[iii] or AQU[iv]). In the past 15 years, universities have begun an evaluation of their services focusing on quality management. On 21 December 2006 the Ministry of Education and Science (MEC) published the *Directrices para la elaboración de títulos universitarios de grado y master* [Directives for the preparation of university degrees and master's degrees]*,* which included the need to introduce a Quality Guarantee System as one of the essential elements in terms of proposing degrees in the future. Until recently, university institutions used an evaluation system based mainly on research activity, leaving the professor's educational duties in the background. That is, the professor was evaluated focusing mainly on the total number of publications he or she produced, the impact factor of where the publication was picked up, and the number of times a specific article was cited. Moreover, teaching evaluation was focused solely on assessing the acquisition of knowledge or lack thereof on the part of the students in terms of higher or lower academic performance (Buendía, 1996). And the students are the ones who have most commonly been chosen to evaluate the quality of teaching performance. In this regard, Tejedor (1996) points out that the reference criterion of the students should not be the only one, since there are a number of contributions to be made in relation to the agents evaluating the professor, with the main ones being the students, other professors, and the professor him or herself (García and Congosto, 2000). In this manner, teaching evaluation has evolved in such a way that it is currently considered to be an extremely important aspect of analyzing the preparatory and professional quality of educational institutions. This new perspective establishes evaluation models based on self-evaluation as a reflexive and participatory process, with the intention of involving all members of the system in order to provide higher quality.

In terms of the instruments used to evaluate teaching performance, a certain disparity has been observed among Spanish universities, although they do have certain aspects in common. A series of studies were carried out in an attempt to analyze the instruments used at each Spanish university. For example, Muñoz, Ríos de Deus, and Abalde (2002) reviewed the instruments and dimensions used at the different universities,

and they designed a new instrument containing 40 items grouped into 10 dimensions (compliance with obligations, infrastructure, curriculum, knowledge-interweaving of the subject matter, methodology, materials, attitude of the professor, evaluation, labs, and satisfaction). Along these same lines, López-Barajas, and Carrascosa (2005) created a questionnaire made up of 25 items broken down into four dimensions: interaction with the students, methodology, teaching obligations and evaluation, means, and resources, for teaching evaluation at the Universidad de Jaén [University of Jaén]. Using the student criterion, they determined which dimensions were most valued and which ones best predicted the teaching assessment, with the "interaction with the students" dimension being the one that most often predicted the overall assessment of the professor. Also, Casero (2008) created an instrument made up of 92 items grouped into infrastructure, curriculum knowledge/interweaving of the subject matter, methodology, materials, attitude of the professor, evaluation, labs, and satisfaction, in order to evaluate university professors, seeking a consensus among students and professors regarding the importance they placed on each of the items, with the goal of creating the ideal instrument for teaching evaluation. González and López (2010) attempted to analyze which teaching competencies Spanish universities consider essential in defining the role of the faculty, and to thereby establish a system of teaching quality evaluation indicators, creating a pilot evaluation instrument made up of 40 items grouped into the following dimensions: design of the curricula/teaching guides for the course, teaching methodology, practical activities, didactic resources, teacher planning, evaluation systems, compliance with teaching obligations, and the attitude of the faculty, where, as opposed to the Casero study (2008), the creation of the faculty competency profile was obtained by means of an expert assessment process and not by consensus among the faculty and students.

Various professors from the Universidad Politécnica and Autónoma de Madrid [Polytechnic and Autonomous University of Madrid] carried out a study in 2010, in which they attempted to develop a system of university teaching quality evaluation indicators based on the analysis of the questionnaires used at different Spanish universities. Although the dimensions proposed by the different universities fluctuate between 2 and 9, the majority of them group the items according to the three dimensions of the model proposed by ANECA (planning, development and result), to which are added other dimensions such as teaching environment or satisfaction. The Universidad de Navarra [University of Navarra] uses a similar questionnaire made up of 24 items grouped into four dimensions (planning, development, results and satisfaction global). Similarly, González and López (2010) carried out a study in which they attempted to estimate the level of validity of the MIECEEES[v] Model, as well as the degree of importance the students placed on each of the items, in order to achieve a quality university. In the end, the instrument was made up of six dimensions, 19 indicators, and 85 items that put into operation the indicators for the quality evaluation model in the European Higher Education Area.

Pozo, Bretones, Martos, and Alonso, (2011) carried out a review, by autonomous community, of Spanish universities that use the DOCENTIA model to evaluate the activity of their university professors, in which they explain the adaptation of this "Docentia-Almería by Pozo, Giménez, and Bretones" model (2009), with the goal of creating a common framework for the evaluation of the faculty at Andalusian universities, establishing

a fourth dimension entitled "innovation and improvement" as the main difference from the ANECA proposal, in which they include elements related to the processes of change incorporated by the professor that may contribute to improved teaching quality.

Although many universities use the same model, there is apparently a great deal of diversity in terms of establishing which factors are representative of a quality teaching methodology at universities. Thus, taking all aspects into consideration, the study objective was to design and validate a measuring instrument allowing for the evaluation of university teaching performance, and to this end we hope to provide validity and reliability to an instrument made up of three dimensions: planning, development and result.

2 METHOD

2.1 Participants

The sample consisted of 1,297 1st to 4th-year university students (762 men and 535 women) aged 18 to 27 ($M = 24.75$; $SD = 5.09$. The original sample was divided into two subsamples. Subsample number 1 consisted of 649 students (410 men and 239 women) and was used to carry out the exploratory factor analysis. Subsample number 2 consisted of 648 students (411 men and 237 women) and was used in the confirmatory factor analysis.

2.2 Measurements

Teaching evaluation. The final version of the *Evaluation of Teaching Performance* questionnaire (CEID) consisted of 28 items which were grouped into three factors: planning, with four items (e.g. "He/she provides clear information about objectives, bibliography, tutorials, contents and assessment methods in the subject's curriculum"), development, with 17 items (e.g. "He/she promotes teamwork") and result, with seven items (e.g. "He/she applies the assessment criteria of the activities as established in the subject's curriculum"). The items were assessed using a Likert scale with a scoring range between 1 (*Completely disagree*) and 5 (*Completely agree*), beginning with the sentence "The professor…".

2.3 Proceeding

The questionnaire was developed based on an extensive review of the literature that provided the most important aspects to be evaluated, which were later reflected in the different items. The drafting of those items emerged from the fieldwork of different university professors who are experts in the subject matter. A first version with 68 items was drafted, which was reduced to 40 items after being reviewed by the experts, and grouped into three factors: planning, development, and result. In addition, the scale was subjected to a general comprehension test, in which the students were asked to read the items carefully and to assess how important they were to them.

The department heads and course coordinators were contacted and informed of the research objective. The questionnaires were completed in the classrooms and before lessons in the presence of the main researcher, while the professor being assessed was not in the classroom. Each course and each professor were assigned an identification number. The main researcher informed the students of the research objective and instructed them on how to properly complete the questionnaire. He/she also resolved any doubts that the students had, and the whole process

required approximately 10 minutes. Participation was voluntary and the participants' anonymity was preserved.

2.4 Data analysis

In order to determine and evaluate the instrument's factor structure, an exploratory factor analysis and a confirmatory factor analysis were carried out, respectively. The instrument's internal consistency was also analyzed using Cronbach's alpha coefficient, and the descriptive statistics (averages and standard deviations) and bivariate correlations of all items were obtained. The SPSS 21.0 and AMOS 21.0 statistical packages were used for the data analysis.

3 RESULTS

3.1 Exploratory factor analysis

After several exploratory factor analyses, certain items were eliminated due to not reaching a minimum rotation of .40. Finally, a factor analysis was carried out for main components with direct oblimin rotation with subsample 1, the results of which were 28 items grouped into three factors: planning (made up of four items: 4, 20, 21 and 22), development (made up of 17 items: 1, 2, 3, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 23 and 26), and result (made up of seven items: 5, 6, 14, 24, 25, 27 and 28). The eigenvalues obtained were greater than 1.00 (1.67, 1.50, and 1.34, respectively), explaining a total variation of 68.2% (38.09%, 15.36% and 14.77%, respectively) (Table 1).

Table 1. Evaluation of Teaching Performance questionnaire (CEID)

Items	Factor 1	Factor 2	Factor 3
1. Presents the minimum content of his/her subject matter, tailored to the students' knowledge	-	.728	-
2. He/She is easily accessible (tutorials, e-mails, etc.)	-	.733	-
3. He/She allows the student to organize and distribute part of the assignments to be performed in the course	-	.736	-
4. He/she provides clear information about objectives, bibliography, tutorials, contents, and assessment methods in the subject's curriculum	.603	-	-
5. He/She informs the students of the competencies they will be expected to acquire	-	-	.569
6. He/She provides me with scientific information that allows me to gain a better and deeper understanding of the subject matter	-	-	.694
7. He/She presents the contents following a clear and logical framework, highlighting the important aspects	-	.605	-
8. He/She allows and encourages student participation	-	-.757	-
9. He/She promotes individual work	-	-.828	-
10. He/she promotes teamwork	-	-.834	-
11. He/She relates the teachings to the professional environment	-	-.577	-
12. He/She provides initial and final overviews of the session and/or subject in class	-	-.632	-
13. He/She encourages student interest and the motivation to learn	-	-.614	-
14. He/She fosters research and a critical spirit in students	-	-	.530

Table 1 (continued)

15. He/She facilitates student-student and student-professor interaction	-	-.712	-
16. He/She attends and responds clearly to questions asked in class	-	.719	-
17. He/She adequately attends to the tutorials requested of him/her	-	.737	-
18. He/She maintains an objective and respectful position with the students	-	.577	-
19. He/She organizes activities for the student to actively participate in course assignments	-	-.653	-
20. He/She designs and relates the classroom content to the lab content	.649	-	-
21. He/She efficiently incorporates and employs ICTs (Information and Communication Technologies)	.498	-	-
22. He/She has a good command of the contents of the course	.692	-	-
23. He/She interweaves the content of the subject matter with other courses	-	.598	
24. He/She applies the established curriculum with a certain amount of flexibility for a better class dynamic	-	-	.625
25. He/She uses material resources that facilitate learning	-	-	.602
26. He/She interacts satisfactorily with the students	-	-.609	-
27. He/She designs the content and develops the course to promote the acquisition of professional competencies	-	-	.591
28. He/she applies the assessment criteria of the activities as established in the subject's curriculum	-	-	.468
Explained variation for each factor	38.09%	15.36%	14.77%
Total explained variation	68.20%		

3.2 Confirmatory factor analysis

A confirmatory factor analysis (CFA) was carried out with subsample 2, based on 28 observed measurements and on the three freely correlating latent constructs (Anderson and Gerbing, 1988). The model was reduced in order to maintain reasonable degrees of freedom (Cecchini, González, & Montero, 2007; Ntoumanis, 2001). Thus, the items that made up each of the dimensions were divided into three groups (Marsh, Richards, Johnson, Roche, & Tremayne, 1994). The validity of the measuring model was considered using a series of fit coefficients, also called goodness of fit measures: χ^2, χ^2/d.f., RMSEA, and the incremental indices (CFI, NFI and TLI). The maximum likelihood estimation method was used along with the *bootstrapping* procedure, since the result of Mardia's multivariate coefficient was 61.56, which indicated a lack of multivariate normality of the data. For this reason, following the example of Finney and DiStefano (2006), the robust maximum likelihood estimation method was used (Byrne, 2001). After an initial analysis, the overall results of the model indicated a moderate fit (χ^2 (59, $N = 649$) = 1899.01, $p = .000$; χ^2/d.f. = 5.47; CFI = .80; NFI = .77; TLI = .81; RSMR = .05; RMSEA = .08). With the goal of improving the fit, we proceeded to carry out an analysis of the Lagrange test (LMtest), which showed that the fit was increased if the errors (9 interactions) of certain items belonging to a single factor, and whose significance level was $p < .000$, were correlated. All items belonging to the development factor. The indices obtained were adequate: χ^2 (69, $N = 649$) = 1051.03, $p = .000$; χ^2/d.f. = 3.12; CFI = .91; NFI = .90; TLI = .90; RSMR = .06; RMSEA = .06.

3.3 Analysis of internal consistency

The internal consistency of each of the factors resulting from the factor analysis (Cronbach's alpha) presented the following results: .70 for planning, .91 for development, and .79 for result. Considering all of the items in a single factor, the consistency obtained was .94.

3.4 Descriptive analysis and bivariate correlations

The average values for the planning, development, and result factors were, 3.95, 3.77, and 3.74, respectively. The correlation analysis data revealed that the three factors were positively correlated (Table 2).

Table 2. Descriptive Statistics and Correlations of the questionnaire's four factors

Variables	*M*	*SD*	1	2	3
1. Planning	3.95	.64	-	.72**	.73**
2. Development	3.77	.57	-	-	.82**
3. Result	3.74	.58	-	-	-

Note: ** $p < .001$; *M* = Mean; *SD* = Standard deviation.

4 DISCUSSION

According to Arranz (2007), society is increasingly demanding that universities operate under criteria of efficacy, efficiency and excellency. So the purpose of this study was to design and validate a measuring instrument that allows us to evaluate university teaching performance. In this sense, it seems important to have processes for evaluating performance methodologies that respond to the educational and professional needs implemented by the EHEA. Thus, in order for these evaluation processes to succeed, it may be essential to prepare the university professor in terms of techniques and procedures for said professor to participate more actively in the evaluation systems. The design and validation of measuring instruments may be a key tool in attempting to optimize teaching performance and thereby carry out an objective assessment of the results (García and Congosto, 2000). The *Evaluation of Teaching Performance* questionnaire was designed for this purpose, and the suggestions contributed by the experts and the respective confirmatory factor analyses showed an instrument made up of 28 items grouped into three factors with adequate internal consistency (planning, development, and result), where teacher planning refers to everything related to the previous process of reflection and design of subject matter (the curriculum, the organization of courses, labs, tutorials, planning of anticipated learning activities, evaluation criteria and methods, teaching materials and resources). The development of the course involves everything related to the execution of and compliance with the curriculum, the teaching and learning activities carried out, anticipated pedagogical activities, as well as applied evaluation procedures. And the results, in terms of educational objectives, refer to the achievements made by the students, and to the aspects involving the revision and improvement of teaching activity, external recognition of teaching duties, and creation of teaching materials. Although we found much diversity in terms of determining which are the main factors to evaluate in teaching practice, the majority of Spanish universities use the model proposed by ANECA as a reference (Pozo, Bretones, Martos, & Alonso, 2011). Along these lines, Pozo et al. (2009) adapt this model for the

Comunidad de Andalucía [Community of Andalusia], grouping the dimensions into planning, development, results, and innovation and improvement. Other universities opt for proposing a higher number of more specific dimensions, such as the design of the curriculum, teaching methodology, practical activities, didactic resources, evaluation systems, and attitude of the faculty (González-López, & López-Cámara, 2010). They also group the dimensions into criteria of interaction with the students, methodology, teaching obligations and evaluation, means, and resources (López-Barajas and Carrascosa, 2005). Casero (2008), in order to evaluate teaching performance, divides the factors into infrastructure, curriculum, knowledge/interweaving of the subject matter, methodology, materials, attitude of the professor, evaluation, labs, and student satisfaction. Whereas González and López (2010) group them into competencies, university information, role of the university student, orientation and tutoring systems, methodology, and student satisfaction. Lukas, Karlos, Etxeberria, and Lizasoain (2014) propose an instrument for evaluating teaching performance made up of 4 dimensions: Teacher planning (5 items), Teaching development and teaching performance (11 items), Result (2 items), and two additional open-ended items (positive aspects of the professor and aspects to improve). Meanwhile, authors such as Bol-Arreba, Sáiz-Manzanares, and Pérez-Mateos (2013) opt for adapting the Student Evaluation of Educational Quality (SEEQ) questionnaire, short version (Marsh, 1987). The newly created version is made up of ten items (Student interest in the course, use of teaching materials, faculty interest in the course, the accessibility of the professor during tutorials and outside the classroom, usefulness of the teaching materials, continuing evaluation tests and their usefulness, the suitability of the evaluation procedures, the quality and explanation of the teaching materials, the professor's motivation for student participation in class, and course workload).

The teaching evaluation carried out by the students ("Students' evaluations of teaching" SET) is used in the majority of universities worldwide as a measurement of teaching quality, so that the results of the assessments provide the professor with feedback for improving his/her teaching (Balam & Shannon, 2010). According to the Kuzmanovic, Savic, Gusavac, Makajic-Nikolic, & Panic study (2013), the indicators of good teaching and quality according to the students are: the availability of the professor, his/her clarity of presentation, the methodological and systematic approach, the information given to the students regarding the coursework, encouraging active participation in class, as well as considering and responding to the students' questions. Similarly, Shukrie (2011) determines that the most effective professors are the ones that know their students' needs and try to increase their motivation, use different teaching strategies, possess good communication skills, and encourage the students to be optimistic about their abilities.

Universities worldwide have developed complex procedures and instruments for evaluating and guaranteeing quality in higher education based on the opinions of their students. This is why we find numerous instruments such as the "Students' Evaluation of Teaching Effectiveness Rating Scale" (Toland, & De Ayala, 2005), the "Teaching Proficiency Item Pool" (Barnes et al., 2008), the questionnaire for student evaluation of teaching (SET 37, Mortelmans, & Spooren, 2009), the Exemplary Teacher Course Questionnaire (Kember, & Leung, 2008), or the "Teaching Behavior Checklist" (Keeley, Furr, & Buskist, 2010).

And although there is a certain variety among them in terms of dimensions or items to evaluate teaching, in general the dimensions identified are similar. Those involving organization, evaluation, and the professor-student relationship seem especially to be the most common in the majority of the instruments. For example, Dodeen (2013) uses a questionnaire with 29 items grouped into five dimensions: Knowledge and organization of the professor (7 items), Clear explanation (6 items), Rating and evaluation (6 items), Teaching methods (4 items), and the Relationship with students (6 items). Toland and Alyala (2005) identify three dimensions: Delivery of course information, professor-student interaction, and the regulation of the students' learning. For Gursoy and Umbreit (2005) there are four dimensions: learning, organization, instruction, and workload.

Although these questionnaires present the validity and reliability necessary for evaluating teaching effectiveness (Khong, 2014; Onwuegbuzie, Daniels, & Collins, 2009), the revision study carried out by Spooren, Brockx and Mortelmans (2013) warns that there are still certain controversies in this regard in the literature, given that the majority of the teaching evaluation questionnaires have not presented sufficient proof of validity.

In the Spanish context, in spite of the diversity of taxonomies regarding the dimensions to evaluate, all of them present various elements in common and have also proven to be valid and reliable instruments for evaluating teaching performance, although an instrument being adequate is a necessary, but not sufficient, condition to guarantee a successful evaluation (Muñiz, & Bartram, 2007).

With the design of this scale it may be possible to have an instrument that allows us to evaluate the teaching performance of university faculty, in such a way that the results obtained in the evaluation may be of great use to the professor, providing objective data. It is therefore essential to consider the aspects related to the process of applying the instrument, such as how the results of said instrument will be used, and it is important that the involved parties receive feedback following the evaluation (Hattie, 2009; Leung, Jiang, & Busserb, 2013; Muñiz, & Fonseca-Pedrero, 2008) and that it be understood as a strategy to improve teaching, involving the students, thereby guiding the teaching activity (Padilla, & Gil, 2008), and that as a result there be implemented a series of plans for improvements and changes to their pedagogical practice, that redefine the teaching-learning process being used up to the present (López Ruiz, 2011), and that they optimize university educational quality, thus providing a more preparatory evaluation (Moore, Walsh, & Rísquez, 2012). But all of this requires a change of attitude, reflected in a transformation of teaching styles (Gargallo, 2008). All of these pedagogical changes involve a great demand on both the professors (Pallisera et al., 2010) and students involved in the process.

5 CONCLUSIONS

After carrying out the different analyses, we have estimated a measuring instrument with adequate psychometric properties that give it validity for the assessment of university teaching performance. The three dimensions evaluated (planning, development, and result) show good internal consistency (.70, .91, .79, respectively), which guarantees the instrument's reliability. However, this study does present certain limitations in terms of the overall character of the evaluation process, which

in this case focuses exclusively on the students' assessment, and although this assessment is necessary, it is not sufficient (Berk, 2013). Other studies are therefore necessary to verify the instrument's factor structure, carrying out an all-encompassing evaluation (Benton, & Cashin's, 2012; Berk, 2009; 2014; Xu, 2012) in which, apart from the student criterion, other assessments are considered, such as faculty self-evaluations, reports from persons outside the institution, work colleagues (Simona, Felicia, & Georgiana, 2014), etc., thereby achieving a more comprehensive teaching evaluation, since this type of evaluation offers valuable information from different perspectives, allowing us to triangulate the information and reorient the professor with the aim of implementing plans for improvement.

Longitudinal studies may also be considered to analyze the validity of this scale's content and construct, expanding the sample of students, and estimating structural regression models corroborating the results obtained.

REFERENCES

Agencia Nacional de Evaluación de la Calidad (ANECA) (2007). *Programa Audit. Guía para el diseño de sistemas de garantía interna de calidad de la formación universitaria*. Retrieved from http://www.aneca.es/var/media/166338/audit_doc01_guidiseno_070621.pdf

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*, 411-423. doi:10.1037/0033-2909.103.3.411

Arranz, P. (2007). *Los sistemas de garantía de calidad en la educación superior en España. Propuesta de un modelo de acreditación para las titulaciones de grado en empresa* (Doctoral dissertation). Retrieved from http://hdl.handle.net/10259/78

Balam, E., & Shannon, D. (2010). Student ratings of college teaching: A comparison of faculty and their students. *Assessment and Evaluation in Higher Education, 35*, 209-221. doi:10.1080/02602930902795901

Barnes, D., Engelland, B., Matherne, C., Martin, W., Orgeron, C., Ring, J., et al. (2008). Developing a psychometrically sound measure of collegiate teaching proficiency. *College Student Journal, 42*, 199-213.

Benton, S. L., & Cashin W. E. (2012). Student ratings of teaching: A summary of research and literature (IDEA Paper no. 50). Manhattan, KS: The IDEA Center. Retrieved from http://www.theideacenter.org/sites/default/files/idea-paper_50.pdf.

Berk, R. A. (2009). Using the 360 degrees multisource feedback model to evaluate teaching and professionalism. *Medical Teacher, 31*(12), 1073-80. doi:10.3109/01421590802572775

Berk, R. A. (2013). Top five flashpoints in the assessment of teaching effectiveness. *Medical Teacher, 35*, 15-26. doi: 10.3109/0142159X.2012.732247

Berk, R. A. (2014). Should Student Outcomes Be Used to Evaluate Teaching? *The Journal of Faculty Development, 28*(2), 87-96.

Bol-Arreba, A., Sáiz-Manzanares, M. C., & Pérez-Mateos, M. (2013). Validación de una encuesta sobre la actividad docente en Educación Superior. *Aula Abierta, 41*(2), 45-54.

Buendía Eisman, L. (coord.) (1996). Evaluación educativa (Monográfico). *Revista de Investigación Educativa, 14*(2).

Byrne, B. M. (2001). *Structural equations modeling with Amos: Basic concepts, application, and programming*. Mahwah, NJ: Erlbaum.

Carreras Barnés, J. (2005). El concepto de competencia y el diseño de planes de estudio a partir de perfiles competenciales. Dos cuestiones que hay que aclarar en el marco de las EEES. In *El Debate sobre las competencias a la enseñanza universitaria* (pp. 8-25). Barcelona: ICE UB.

Casero, A. (2008). Propuesta de un cuestionario de evaluación de la calidad docente universitaria consensuado entre alumnos y profesores. *Revista de Investigación Educativa, 26*, 25-44.

Cecchini, J. A., González, C., & Montero, J. (2007). Participación en el deporte y fairplay [Participation in sport and fair play]. *Psicothema, 19*, 57-64.

Dodeen, H. (2013). Validity, Reliability, and Potential Bias of Short Forms of Students' Evaluation of Teaching: The Case of UAE University. *Educational Assessment, 18*(4), 235-250. doi:10.1080/10627197.2013.846670

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modelin. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: a second course* (pp. 269-314). Greenwich: Information Age Publishing.

Garcia Ramos, J. M., & Congosto, E. (2000). Evaluación y Calidad del Profesorado. In T. González Ramírez (Ed.), *Evaluación y Gestión de la Calidad Educativa. Un Enfoque Metodológico* (pp. 127-157). Málaga: Aljibe.

Gargallo, B. (2008). Estilos de docencia y evaluación de los profesores universitarios y su influencia sobre los modos de aprendizaje. *Revista Española de Pedagogía, 66*(241), 425-446.

González, I., & López, I. (2010). Validación y propuesta de un modelo de indicadores de evaluación de la calidad en el la universidad. *Revista Iberoamericana de Educación, 53*(6), 1-13.

González López, I., & López Cámara, A. (2010). Sentando las bases para la construcción de un modelo de evaluación a las competencias docentes del profesorado universitario. *Revista de Investigación Educativa, 28*(2), 403-423.

Gursoy, D., & Umbreit, W. T. (2005). Exploring students' evaluations of teaching effectiveness: What factors are important? *Journal of Hospitality and Tourism Research, 29*(1), 91-109. doi:10.1177/1096348004268197

Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on archievement*. Oxford, UK: Routledge.

Keeley, J., Furr, R. M., & Buskist, W. (2010). Differentiating psychology students' perceptions of teachers using the Teacher Behavior Checklist. *Teaching of Psychology, 37*, 16-20. doi:10.1080/00986280903426282

Kember, D., & Leung, D. (2008). Establishing the validity and reliability of course evaluation questionnaires. *Assessment & Evaluation in Higher Education, 33*, 341-353. doi:10.1080/02602930701563070

Khong, T. L. (2104). The Validity and Reliability of the Student Evaluation of Teaching: A case in a Private Higher Educational Institution in Malaysia. *International Journal for Innovation Education and Research, 2*(9), 57-63.

Kuzmanovic, M., Savic, G., Gusavac, B. A., Makajic-Nikolic, D., & Panic, B. (2013). A Conjoint-based approach to student evaluations of teaching performance. *Expert Systems with Applications, 40*(10), 4083-4089. doi:10.1016/j.esw.2013.01.039

Leary, H., Walker, A., Shelton, B. E., & Harrison, F. M. (2013). Exploring the Relationships Between Tutor Background, Tutor Training, and Student Learning: A Problem-based Learning Meta-Analysis. *Interdisciplinary Journal of Problem-based Learning, 7*(1), 40-66. doi:10.7771/1541-5015.1331

Leunga, X. Y., Jiangb, L., & Busserb, J. (2013). Online student evaluations of hospitality professors: A cross-cultural comparison. *Journal of Hospitality, Leisure, Sport & Tourism Education, 12*(1), 36-46. doi: 10.1016/j.jhlste.2012.10.001

López-Barajas, D., & Carrascosa, J. (2005). La evaluación de la docencia universitaria. Dimensiones y variables más relevantes. *Revista de Investigación Educativa, 23*(1), 57-84.

López Ruiz, J. I. (2011). Un giro copernicano en la enseñanza universitaria: formación por competencias. *Revista de Educación, 356*, 279-301. doi:10-4438/1988-592X-RE-2010-356-040

Lukas, J. F., Santiago, K., Etxeberria, J., & Lizasoain, L. (2014). Adapting to the European Higher Education Area a questionnaire on student opinion about the teaching of lecturers. *Revista Electrónica de Investigación y Evaluación Educativa, 20*(1). doi:10.7203/relieve.20.1.3812

Marqués, M. (2008, September). Una experiencia madura de innovación educativa. In *Jornadas Nacionales sobre Estudios Universitarios de los Proyectos de Convergencia a la Realidad de los Nuevos Títulos.* Universidad Jaume I.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253-288. doi:10.1016/0883-0355(87)90001-2

Marsh, H. W., Richards, G. E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical self-description questionnaire: psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology, 16*, 270-305.

Moore, S., Walsh, G., & Risquez, A. (2012). *Estrategias eficaces para enseñar en la Universidad.* Madrid: Narcea.

Mortelmans, D., & Spooren, P. (2009). A revalidation of the SET37-questionnaire for student evaluations of teaching. *Educational Studies, 35*, 547-552. doi:10.1080/03055690902880299

Muñiz, J., & Bartram, D. (2007). Improving international tests and testing. *European Psychologist, 12*, 206-219. doi:10.1027/1016-9040.12.3.206

Muñiz. J., & Fonseca-Pedrero, E. (2008). Construcción de instrumentos de medida para la evaluación universitaria. *Revista de Investigación en Educación, 5*, 13-25.

Muñoz Cantero, J. M., Ríos de Deus, M. P., & Abalde, E. (2002). Evaluación Docente vs. Evaluación de la Calidad. (RELIEVE). *Revista Electrónica de Investigación y Evaluación Educativa, 8*(2), 103-134.

Ntoumanis, N. (2001). A self-determination approach to the understanding of motivation in physical education. *British Journal of Educational Psychology, 71*, 225-242. doi:10.1348/000709901158497

Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. T. (2009). A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity, 43*, 197-209. doi:10.1007/s11135-007-9112-4

Padilla, M. T., & Gil, J. (2008). La evaluación orientada al aprendizaje en la Educación Superior: condiciones y estrategias para su aplicación en la docencia universitaria. *Revista Española de Pedagogía, 66*(241), 467-486.

Pallisera, M., Fullana, J., Planas, A., & del Valle, A. (2010). La adaptación al espacio europeo de educación superior en España. Los cambios/retos que implica la enseñanza basada en competencias y orientaciones para responder a ellos. *Revista Iberoamericana de Educación, 52*(4), 1-13.

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. *Educación XXI, 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.

Salmerón, L. (2013). Actividades que promueven la transferencia de los aprendizajes: una revisión de la literatura. *Revista de Educación*, special issue 2013. doi:10.4438/1988-592X-RE-2013- EXT-253

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research, 83*(4), 598-642. doi:10.3102/0034654313496870

Shukrie, R. (2011). Professor's performance for effective teaching (Kosovo case). *Procedia Social and Behavioral Sciences, 12*, 117-121. doi:10.1016/j.sbspro.2011.02.015

Simona M., Felicia, S., & Georgiana, C. V. (2014). Content analysis regarding the multiple evaluation of the academic activity in CMU. *Global Journal of Guidance and Counselling*, *4*(1), 25-31. Retrieved from http://www.world-education-center.org/index.php/gjgc/article/view/3196

Tejedor, F. J., & García Valcárcel, A. (1996). Evaluación institucional en la Universidad. *Revista Galega de Pisopedagoxia, 6*, 101-146.

Toland, M. D., & De Ayala, R. J. (2005). A multilevel factor analysis of students' evaluations of teaching. *Educational and Psychological Measurement, 65*, 272-296. doi:10.1177/0013164404268667

Vansteenkiste, M., Soenens, B., Sierens, E., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology, 101*, 671-688. doi:10.1037/a0015083

Xu, Y. (2012). Developing a comprehensive teaching evaluation system for foundation courses with enhanced validity and reliability. *Educational Technology Research and Development, 60*(5), 821-837. doi:10.1007/s11423-012-9240-y

i ANECA: Agencia Nacional de Evaluación de la Calidad y Acreditación [National Agency for the Evaluation of Quality and Accreditation]

ii ACSUG: Axencia para a Calidade do Sistema Universitario de Galicia [Agency for the Quality of the University System of Galicia]

iii AVAP: Agencia Valenciana d'Avaluació i Prospectiva [Valencian Agency of Evaluation and Planning]

iv AQU: Agència per a la Qualitat del Sistema Universitari de Catalunya [Agency for the Quality of the University System of Catalonia]

v MIECEEES: Modelo de Indicadores de evaluación de la calidad en el Espacio Europeo de Education Superior [Model of Indicators for the evaluation of quality in the European Higher Education Area]. Financed by the Ministry of Education and Science, General Directorate of Universities. Code: EA2006-0116