GENERAL SECTION — ARTICLES

ICT and 360º evaluation: Improving professional skills in higher education in Spain

Daniel David Martínez-Romera and Sara Cortés-Dumont

ICT and 360º evaluation: Improving professional skills in higher education in Spain

Daniel David Martínez-Romera and Sara Cortés-Dumont*

Abstract: The current dynamics of knowledge and innovation generation are faced with, sometimes, incompatible social and cultural trends. Something to which the University is not oblivious. Based on contemporary studies and our own experience, one of the clearest tensions has to do with the ability to judging based on reasons and not emotions. To help with this, in the educational context, the 360-degree evaluation can be a useful instrument in terms of the strengthening of their objective judgment. It is a technique that was conceived to exercise objective evaluation, in several and concurrent ways, which also allows to know their degree of objectivity. To demonstrate its potential, a sample of 56 students was used, taking the teacher's grading as a reference. The task involved self, peer and inter-group assessment. The methodology was mixed, with support in descriptive statistics for quantitative grades and natural language processing for comments and clarifications. It was possible to detect differences in behavior depending on the type of analysis (self-assessment, peers, and groups), as well as determining which students were most qualified to assess objectively. Another issue was the general reluctance to explain numerical grades through notes. Here we consider several factors to explore, among which we highlight: the digital tool used and the time available; the phase of the academic course and the characteristics of the task; and subjective bias, especially in cases of low score. About the experience, it has also been possible to detect some difficulties derived from the application of a methodology, as complex and demanding as the one

* **Dr. Daniel David Martinez-Romera** (corresponding author, ddmartinez@uma.es), PhD in Education and PhD in Geography, is Associate Professor in the Faculty of Education at the University of Málaga, Spain.
  Mtr. Sara Cortés-Dumont (scortes@ujaen.es), Master in Geographic Information Systems, is Postgraduate Teaching Assistant in Geography in the Faculty of Humanities at the University of Jaén, Spain.
  More information about the authors is available at the end of this article.

used, which generated more than 3000 evaluations. In any case, and in view of the data obtained, we consider that the results corroborate the practical utility of this approach and invite to explore additional aspects within this area.

Keywords: educational technology; formative evaluation; educational research; statistical analysis; comparative analysis.

I. Introduction

The main raison d'être of university educational innovation is to improve the training of new generations of professionals. This can be approached by many different angles. One of them is the ability to improve and reflect the development level of the professional skills that are to be acquired, such as the ability to deliver a fair judgment.

For this to happen, people must first of all become aware of the subjective pressure their decisions are subject to on a daily basis. This is a relevant issue in terms of assessment (Arreola Rico 2019; Curcu 2008; Galaz and Toro Arévalo 2019; Jornet Meliá et al. 2020; Oliveras Boté 2019), which goes beyond this field, extends into society (González-Such et al. 2021; Jornet Meliá et al. 2011; López Aguilar et al. 2020) and is even addressed in educational policies (Álvarez-López and Matarranz 2020; Elías 2017; Perales-Montolío et. al. 2014).

The normative solution reached by the Spanish educational system has been clear and widely agreed upon. The training of new generations of professional is not conceivable without first specifying what is expected of them in terms of the competence, skills, and capabilities they need to have, as well as the plans and programs that make this possible (Ortiz-Revilla et al. 2021; Rodríguez-Gómez et al. 2017; Sarceda-Gorgoso and Rodicio-García 2018). The ability to evaluate impartially, formulated in different ways, either on an individual or on an integrated basis and often associated with critical thinking, is always present (Vendrell i Morancho and Rodríguez Mantilla 2020).

Even taking into account the exceptional situation caused by COVID-19, it is clear that the role of educational technologies has become more and more relevant in all aspects of the training process. Both in university and pre-university environments, especially in the faculty of Education and in relation to specific didactic methods (Arancibia Herrera 2016; Martínez Romera 2019; Martínez Romera et al. 2020; Urquidi Martín et al. 2019).

Teaching social contents is, in this case, unavoidable, both for their formative relevance and because of the central position occupied by their objectivity-subjectivity pressure, due to the presence of the ideological and

emotional dimensions (in a broad sense), even in the university (Furedi 2018). This could affect the evaluation on both the form but also on the subject matter: the former being the preparation and exposition of a topic and the latter the interpretations and approaches on the mention topic.

People are more likely to accept a natural or mathematical fact -even when demonstrated by someone with whom there is an ideological disagreement- than a social or cultural fact, assuming that both facts are rigorously true. Thus, the subjective bias is not limited to the fact that there is a preference on how a topic has been presented, but there is also a judgment of the (or from the) ideology of the actors involved, while the evidence, the facts, and their methodological structure are left in the background. Thus, learning how to professionally manage form vs. matter, necessarily requires taking those relationships established between evidence, preconceptions, and subjectivity into consideration.

The role of information technology (ICT) can be critical in this regard, insofar as it can facilitate issues that otherwise could not have been addressed due to physical (availability of adequate space) or demographic (overcrowded classrooms) constraints. Digital rubrics (Cebrián-de-la-Serna and Bergman 2014; Fernández-Quero 2021; Ferreiro Concepción and Fernández Medina 2020; Grande de Prado et al. 2021), have proven very useful to solve problems such as attendance and participations. Thanks to them, it is possible to establish an evaluation tool that can be applied in multiple ways to establish objective judgments about curricular work.

One of the most interesting, but also the most complex to implement, is the 360-degree assessment (US Office of Personnel Management 1997). It has its origin in Germany in the early 1930s, as an instrument to improve the selection of military officers. After World War II, its use became widespread in the business world, as a tool for improving the selection of candidates. Despite the good results it offers, its use has not been too abundant in the analog era due to the complexity of its procedure (a lot of evaluation forms are involved). It has been with the arrival of the computer revolution, especially Internet, that its use has been increasing in number and scopes.

This technique is characterized by the use of three points of view that converge on the same issue, fact or activity. In general terms, a previously designed questionnaire is fulfilled recursively for comparative purposes: the vision of the person who makes or performs the object of evaluation; the vision of the peers; and the vision of the professional evaluator. This is done for each of the participants, hence the circular procedure that gives it its name. Upon completion, a significant volume of information has been generated, that can be analyzed from various angles. This is intended to

obtain a more contrasted and detailed judgement, which can overcome the subjective biases inherent to the human being.

Adapting this to the educational context, we have that the introduction of both self-assessment and peer evaluation, in addition to the teacher's, means involving students in decision-making about content and/or the performance of third parties. This practice is very interesting for the educational field (Báez-Rojas et al. 2021; Barba Aragón 2020; Dagal and Zembat 2017; Liu et al. 2021; Martínez Romera 2017; Meghdad et al. 2020), especially for social sciences and how they are taught.

In line with all the above, the following research question will be addressed: Is it possible to use the 360-degree technique in digital contexts to strengthen student impartial judgement? Two subquestions derive from it: Is it possible to identify and improve the skills of individuals? What role does the working group, and the class-group, play in the individual bias?

To carry out the experience and collect the necessary data, the CoRubric web application was used to design the questionnaire. A free tool for the design, application and analysis of evaluation rubrics in digital contexts, developed by Daniel Cebrián Robles (2019) and GTEA, a research group in educational technology from the University of Málaga (Spain).

II. Methodology

A case study was presented based on 56 bilingual (Spanish-English) primary school teacher training. The project used a digital assessment rubric based on the 360-degree model built using CoRubric. The presentations of the final classwork carried out in small groups (15) were evaluated in groups of 4 to 6 students. They consisted of the complete development of an educational activity (didactic unit) on the curricular contents of Social Sciences in Primary Education.

The instrument was developed and validated according to the standards in use (Cubillos-Veja and Ferrán-Aranaz 2018; García-Valcárcel Muñoz-Repiso et al. 2020; López-de-Arana Prado et al, 2019; Ortega-Quevedo et al. 2020; Tejada-Fernández et al. 2015; Usart Rodríguez et al. 2020). Both the form and the subject matter of the object to be evaluated were considered, a set of dimensions and criteria were established halfway between generalization and detail, and measurement scales were created according to each case, duly provided with semantic content. Before putting it into practice, this first draft was submitted and adjusted according to the opinion of experts; after this, a first pilot adaptation test was carried out, which also served to establish the mechanics of use, and no changes were needed.

Following the nomenclature (Gatica-Lara and Uribarren-Berrueta 2013; Pozuelos Estrada et al. 2020) for analytical rubrics, the tool covered 4 concepts and 9 unweighted aspects to be evaluated. The measurement scale, a closed Likert-type scale, presented four levels in 8 cases and three in 1. Each and every one of these levels has evidence descriptors (graded semantic content) to determine the level of compliance. The result of the rubric generates a final numerical score ranging from 0 to 100. Table 1 shows the final structure of the implemented tool.

Table 1

General layout of the assessment rubric

1. Content structure
1.1. Information quantity and relevance*
1. It has conceptual errors and focuses on ancillary issues
2. It has conceptual gaps and tends to focus on ancillary issues
3. It has some inaccuracies and focuses on the main topic of the work
4. It addresses all the theoretical aspects satisfactorily and is well focused on the subject
1.2. Degree of structuring
2. Communication with the audience
2.1. Oral communication
2.2. Body language
2.3. Management of the resources used
3. Interaction with the audience
3.1. Ability to motivate and create interest
3.2. Control of interactions
4. Use of technological resources
4.1. Quality of the resources used
4.2. Formal aspects of the presentation

* Evidence for each of the levels, it is presented as a sample in order to keep a compact view of the tool.

Source: compiled by author.

The digitalization and student use of the rubric was carried out using CoRubric. Fifteen additional assessment objects, the workgroups, were created, so that 56 evaluators had to assess 70 evaluation points using the 360-degree logic: self-assessment, peer evaluation (and workgroup evaluation) and teacher evaluation. The process generated 3042 assessments, 77.60% of the planned number of assessments (3920).

The analysis used mixed methods, consisting of the conversion of the qualitative assessments into discrete quantitative categories and the consideration of the (optional) observations that each aspect allows to be made in addition to its assessment. This latter aspect proved to be testimonial, with 34 annotations. Personal data anonymity was guaranteed by assigning a numerical identifier (IDx) to each participant.

The results were analyzed with SPSS v.26, and Pandas and scikit-learn (Pedregosa et al. 2011) were used for the advanced matrix manipulation with automata. For qualitative analysis, annotations comments, natural language processing (Hussen et al. 2021) was used, through NLTK (*Natural Language Toolkit*) library in Python.

II.1. Data control

The last methodological issue will be the first to be addressed in the analysis. Since its robustness depends on whether it can be continued. Two aspects must be checked here: that the data obtained is valid and reliable.

The question of validity has to do with the correct use of the evaluation scale assigned to each item. Since a digital application is used, this aspect is well controlled and it is not possible to get answers out of range. An advantage derived from the use of ICT support over the free manipulation of physical questionnaires by the participants. Therefore, this aspect will have no impact in our case, but it was necessary to point it out.

Regarding the reliability of data, we must rely on descriptive statistics. This type of control allows us to rule out that they are meaningless. In terms applied to our experience, for example, this would happen if everyone responded with a constant score to all items. Or if they do it through clear patterns, like: minimum/maximum grades, in an alternate way; or counter loop, meaning increase gradually the response value for each successive item until max-range is reached, and start over (or vice versa).

We find here Cronbach's Alpha coefficient as a consolidated reference (Barbera et al. 2021; Bujang et al. 2018; Emerson 2019). From its application a quotient is obtained that allows us to have greater certainty of the reliability of the dataset. Then a second analysis, called item-total correlation, must be

performed to ensure that each item is also consistent. If not, two types of action are derived: modify or delete it. The first leads to reformulating it and answering it again. If not possible, then it should be removed and the data reliability recalculated for the remaining set. This aspect is central to our analysis, so it will be the first stop of it.

III. Analysis

III.1. *Tool validation*

The scale reliability analysis applied to the aspects yielded a Cronbach's Alpha of 0.791, 0.795 with standardized items, above the 0.7 value considered the minimum necessary for tools designed for individual analysis in education (Taber 2018). The item-total correlation was above 0.3 in all cases, which meant an individual Cronbach's Alpha always higher than 0.75, thus reformulating or discarding any aspect was not necessary. (Frías-Navarro and Pascual-Soler 2021). In view of the above, the instrument exceeded the reliability requirements necessary to develop the data analysis.

III.2. *Self-evaluation*

Forty-five of the fifty-six participants completed the self-evaluation (80.4%). The average grade was 92.12 out of 100, with a standard deviation of 15.54 points. The extreme situations ranged from 0, reported by one individual (ID40), to 100, reported by 5 people (ID4, 5, 10, 22 and 24). Taking the teacher's grading as a reference, this had an average value of 84.57, 7.55 points lower than the student grade, and showed a similar standard deviation (14.0 points). ID40 and ID5 had a double self-evaluation, with average grades of 97.33 and 94.0. This meant that ID9 and 47 had the lowest self-evaluations (71.78 points). Figure 1 shows the comparative behavior.

Almost all of the grades were above the 50-point range in both cases. The student perception of their own performance was systematically higher and more concentrated than that observed by the teacher. The compared analysis made it possible to establish some behavioural patterns:

i. No significant discrepancies could be observed in the first section, up to approximately 40 points.
ii. between 40 and 60 points, there were specific situations with a higher grade from the self-evaluation than from the teacher.

iii. Most students were assigned between 60 and 90 points by the teacher, a discrepancy that peaked at 85 points, where the density of students quadruples that found in the self-assessments.

iv. At 90 points and above, the scenario was reversed, with the teacher finding fewer cases than those reported by the self-assessments, a situation that reached its turning point at 93 points, where the rate given by the teacher is one third of the self-assessed.

Therefore, the demographic grading patterns show significant discrepancies in favour of self-perception as compared to the teacher's evaluation. Only 5 people were stricter with themselves than the teacher, with ID47 being the strictest with a difference of 9.32 points (71.78 vs 81.1, respectively). While 40 students graded their work higher, to varying degrees: 23 by up to 10 points in their favour, 14 by up to 20 and 3 by more than 20; with ID42 reaching a record of 24.13 points (97.33 vs. 73.2).
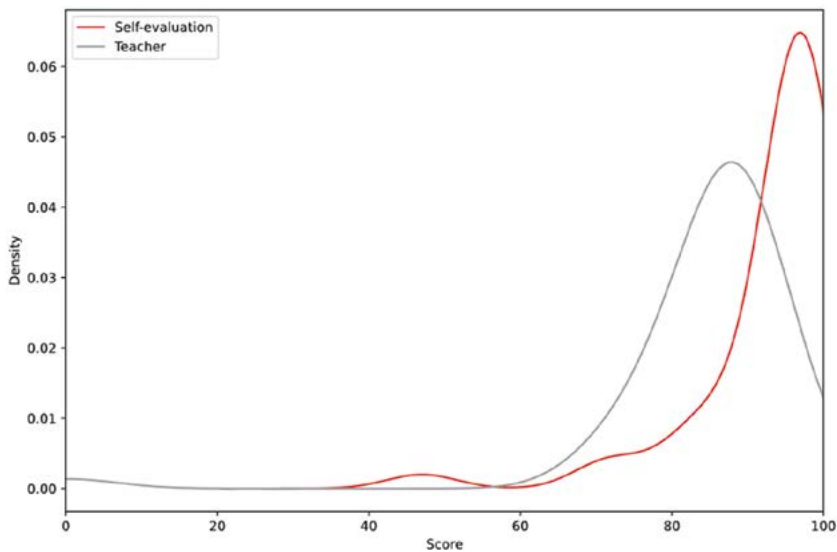

Figure 1

Comparative distribution of self-evaluation and teacher grades

III.3. *Hetero-assessments*

Fifty-six students participated in the peer evaluation, generating a total of 2,384 records, excluding the workgroup evaluations. The minimum grade is

5, given by ID27, and the maximum is 61, by ID18 and ID50. All cases above 55 are the result of the re-assessment by one or more peers. To analyze the behavior of the evaluators, a comparison was made with the teacher in terms of the average grade given in each case, as shown in Figure 2.
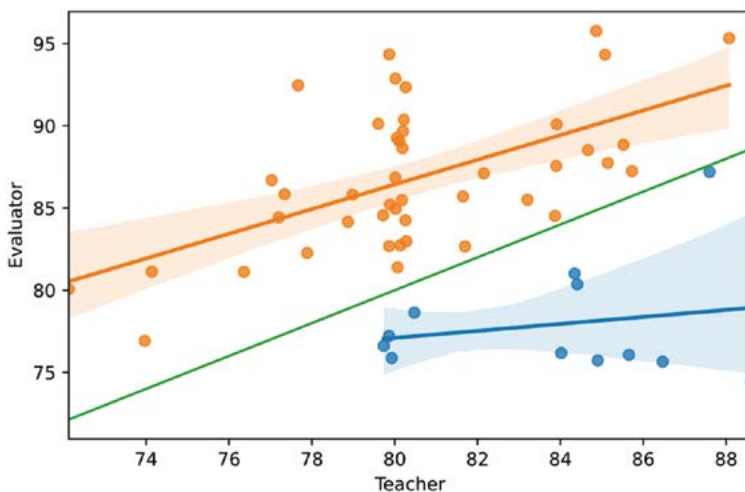

Figure 2

Comparison of the evaluations made, in average value

The diagonal dividing the graph into two areas represents the ideal line of perfect correspondence between the teacher's grade and the grade given by the students. Thus, the light dots at the top represent evaluators who gave higher grades than the teacher, and the dark dots at the bottom represent evaluators who were harsher than the teacher. The first group comprised of 44 students (78.57%) and the second of 12 (21.43%).

In the group of overgrades, in absolute terms ID30 stands out as the most extravagant value, with an average score of 95.77 points compared to 84.86 given by the teacher; the lowest case is represented by ID21 with a value of 76.92 points compared to 73.97 by the teacher. In relative terms, the highest discrepancy range belonged to ID10, who gave an average grade of 92.46 compared to 77.67 by the teacher, which implies a discrepancy of 14.79 points; while the most similar is ID26 with a grade of 84.52 compared to 83.87 by the teacher, which represents a deviation of 0.65 points.

In the group of undergrades, in absolute terms ID52 stands out as the most extravagant value, with an average score of 87.19 points compared to

87.6 given by the teacher; the strictest student was ID27 with 74.44 points compared to 88.61 by the teacher. In relative terms, the largest discrepancy was found in ID27, with 14.17 points of difference, thus coinciding with the lowest absolute extreme of the subgroup and of the whole class; and the most convergent coincides, similarly with ID52, with a discrepancy of 0.41 points. The standard deviation of the grades is 5.52 points for the students and 3.51 for the teacher; by subgroups, the overgraded subgroup yields a value of 4.31 and 3.28 and the undergraded subgroup a value of 3.53 and 3.14, respectively.

In addition, the regression curves and the estimated confidence interval range are also shown. There is a clear discrepancy in the slope between the two, so that the slope of the undergrades is almost horizontal, which describes a much more constant behavior of the criteria issued, with a marked widening of the confidence intervals towards the maximum values. The overgrade curve is slightly less steep than the ideal diagonal, indicating that the convergence of values occurs towards the upper end, the opposite of what happens in the undergrades. The confidence interval is more homogeneous, with a tendency to widen at the outer edges.

III.4. Group assessment

A second approach to hetero-assessment was taken into consideration, focusing the analysis on the evaluations issued on each work group as a singular entity, and it was also compared with the teacher's criteria based on the 613 grades given. The average number of grading per group was 40.87, with a minimum of 31 (J), a maximum of 49 (H), and a standard deviation of 6.83. Figure 3 shows the results of this comparison, arranged in descending order according to the teacher's grading.

There are no analogies between student and teacher behaviour. In contrast, there are some marked discrepancies, especially those affecting Ñ, J and I, in the group of those undergraded by the teacher, coinciding with relative minimums, and absolute in the case of J, which is also the group that has received the fewest evaluations; M, G, B and A show the highest overgrades, coinciding with relative maximums, and absolute in the case of M. The group with the most evaluations issued, H, is in a context of inflection in the relative maximum in a section of clear overgrade. The highest student/ teacher convergence was observed in N, D and K, with no additional evidence to establish further relationships.

The groups' internal behaviour was highly heterogeneous, especially at the lower end. Within the 51 student grades with a fail grade(<50), 9 out of the 15 groups are identified: A (6), C(1), D(7), E(7), G(4), H(7), I(2), J(17) y
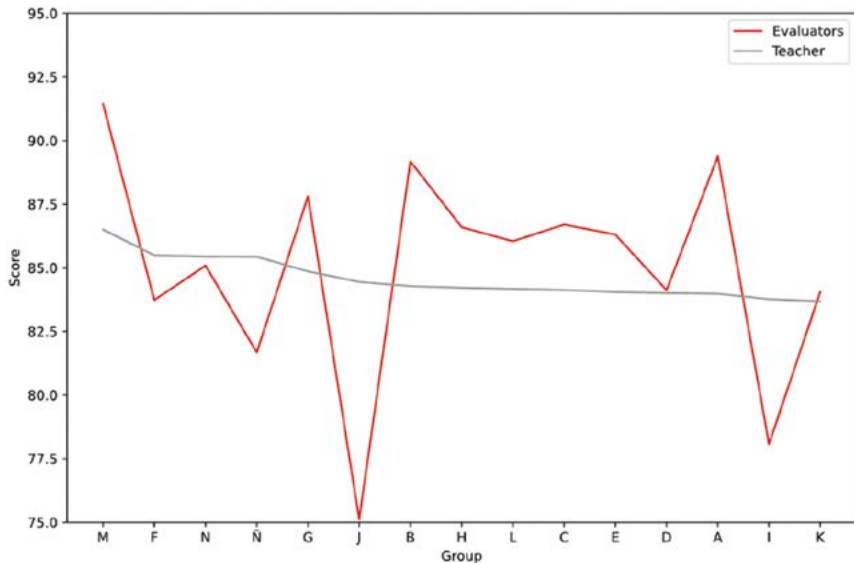
Figure 3

Comparative grading of workgroups, sorted by teacher's grading

N(2). One third of all the lowest grades fall on J. In addition, group grades compared to the average of its members are also lower. In statistical terms, there is a discrepancy in the criteria for grading the group as a unit rather than as a group of individuals.

III.5. Notes

Comments regarding the grades were testimonial and when they had a useful semantic value it was to indicate a low grade. Thus, ID11 justified his/her zero for evidence 2.1 to ID31 on the fact that she did not actively participate ("*she didn't speak*"), and the same happened to ID22. An exception to this pattern is ID8, who justified a high grade in this same item for ID0 by giving her added merit due to her circumstance ("*She has presented her part without having the best tone of voice, so her effort must be congratulated*").

The above comments were always about people from different groups. But there are also redundant situations, as in the case of evidence 3.1. ID45 commented on ID47 -belonging to the same group (M)- to whom he/she awarded 80.67 points, arguing that "*she made the presentation of the project as if the participation of the audience was not necessary*."

A third situation was detected whereby every member of a group had the same traits as those of the group itself. ID21, in relation to evidence 3.2, assigned to all members and to the group itself, a score below 30 in all cases, arguing that "*they did not provide any time for questions or doubts about their project*", a comment that is noted for ID35, 37, 42 and the group itself (F), apart from ID36, also a member of this group.

Apart from the above, no elements were qualitative interesting as they did not significantly effect the assessment: ID22 pointed out that on evidence 4.2, group B exceeded the established time ("*They finished late*"), a comment which resulted in a grade of 90.67 points; several cases are also identified in which the comment reflects the chosen grade (1, 2, 3, 4).

III.6. *The influence of gender*

Although the sample had a clear female majority, 44 of the 56 participants were females, the high number of evaluations generated by each student allowed to consider the average grades given as an exercise in behavioral exploration in terms of peer judgment. The main results are laid out in Table 2.

Table 2

Mean ratings between sexes and corresponding teacher ratings

Gender of evaluator	Gender rated		Teacher	
	M	F	M	F
M	88.39	85.24	79.72	82.81
F	87.03	84.86	79.95	83.01
Mean	87.36	84.95	79.89	82.96

Source: compiled by the author.

Apart from the considerations on the grades in absolute terms which are similar to the student/teacher discrepancies identified above, it is possible to point out an additional new aspect associated with the discriminatory analysis by gender. On average, the grades given between males were more than 4 points higher than the average grade between females, a situation that was almost the opposite when viewing the grades given by the teachers. In the case of female-to-male and male-to-female grades, the difference is less than 2 points, and less than 3 in the case of the teacher.

IV. Discussion and conclusions

Before delving into the comment on the ins and outs of the research, we consider it necessary to contrast the results with the sources presented in the introduction from a global perspective. In order to establish the possibilities of use in other areas and places, as well as to determine to what extent the experience is generalizable. To do this, we will confront the topics addressed in the introduction that have served as theoretical support: the subjectivity of judgment in the university context; the identification of formal constraints for its reduction; the uniqueness of the social sciences as an obstacle; and the general difficulties of technological implementation. After this, the discussion of results will be developed and, finally, as a closing syllogism, the considerations related to the research question, and the derived sub-questions, will be addressed in the final paragraphs.

The presence of subjectivity in the evaluation has been verified, something that must be assumed as part of it and the free will of the students as a social subject (Curcu 2008, 212-213). Our position here is to agree on the diagnosis but disagree on the consequence, since today it is the subjective component that has been oversized and that needs to be reweighted, in our geographical context, as will be discussed later. Arreola Rico (2019), considers that this must be extended not only to teachers in training, but also to teachers in practice, because emotional competence must be present in the evaluation process. We understand that this could not be otherwise but, as in the previous case, we offer a different angle: today, the ability to discern and assume facts from personal interest is under question. This is an inalienable teaching skill in our opinion.

It still remains to be addressed, at this point, that bias may be present even in the teacher. As Galaz and Toro Arévalo (2019, 374) point out, evaluation expresses a set of relationships and spaces loaded with meaning for those involved. What is extended not only to his teaching, but also to his research, as demonstrated by Oliveras Boté thesis (2018). In both cases, the confrontation observed here between objectivity and subjectivity is reflected on the general theoretical plane, as a reality on which to act.

We opted on this occasion for the wake-up call of the later, since problems such as confirmation bias find in subjectivity fertilizer for its growth. This is not posed as an exclusionary dichotomy, but as a complex dialectic. Emotion is a source of implication and motivation of first order, as González-Such et al. (2021, 292) recalls from the perspective of social cohesion. With López Aguilar et al. (2020, 99), all the above makes it crucial to apply these findings in the evaluation processes and establish greater

communication between the educational research carried out and teaching practice, something about which there is still much to do.

A legal framework, the Spanish one, has been assumed for the development of the experience, this introduces conditioning factors. Broadly speaking, we can establish the existence of two educational models in the world: those based on competences and those based on content. Historically, all countries were in the second model, but since the last third of the twentieth century a change towards the first, consolidated at the beginning of the twentieth century, developed in some countries. Spain is immersed in a process of European convergence (European Higher Education Space), which is trying to generalize beyond its borders with the inclusion of countries such as Russia and Turkey. This is a competency-based proposal.

It is clear that the experiences indicated here, such as that of Sarceda-Gorgoso and Rodicio-García (2018), only have applicability under this paradigm. In our opinion, oriental models, such as Chinese, Korean or Japanese, clearly oriented towards content, present different problems. In this case, their concern is to promote the emotional involvement of university students to improve their involvement and performance (Shafait et al. 2021; Zhoc et al. 2020), as they belong to a sociocultural context where emotions tend to be more restrained.

The characteristics of the social sciences, as a subject of work, must be considered with respect to the natural or experimental sciences. Clearly, objectivity is much more assumed in the latter as an intrinsic characteristic. But it comes into conflict with ideology and emotionality, understood here in a broad sense, when we address the former. This fact is generalizable beyond this experience, because it connects with intrinsic characteristics of the different knowledges.

Thus, from the calm reading of the arguments of Furedi (2018), a clear conclusion is obtained: the vast majority of the academic scandals pointed out that lead to the dismissal of professors, to the censorship of research, activities and arguments, and even to the promotion of self-censorship as a survival strategy, occur in the field of social sciences. We consider the above as further evidence of the need to work, precisely, on the strengthening of rational thought in the branches of knowledge that in the university are being more gripped by subjectivism and postmodernism (Andrade 2019; Kestel and Korkmaz 2019; Milliken 2007; Searle 1995).

Undoubtedly, the technological perspective is the most positive when assessing the possibilities of implementation and generalization of this type of strategies, for several reasons. First, because the computer support required is minimal; any computer equipment capable of surfing the Internet

is sufficient, including mobile devices, which relaxes the pressure on available resources in contexts and countries with limited resources (Das 2019; Jogezai et al. 2020). Second, the virtualization of experience opens new windows of flexibility on synchronous and diachronic evaluation processes, as pointed out by the aforementioned Grande Prado team (2021, 53-54). And third, we must not forget that this type of technology, of which the new generations are native, is associated with an increase in participation inside (Zweekhorst and Maas 2015) and outside the classroom (Mano 2021).

About results discussion, the analysis of self-assessments is particularly interesting in the field of acquisition of professional competencies. This is because making an impartial judgment on one's own abilities, skills and performance is subject to the greatest possible pressure considering one's own interests. This ability to assess based on and adjusted to reality requires a habit that needs to be developed and strengthened. However, even in the field of education degrees, as is the case here, this type of assessment is often displaced by the more usual peer evaluation.

It is worth asking about the causes that lead to giving these clearly discrepant evaluations. Based on the authors' experience and the experiences analyzed in the initial sections, the authors believe that three of them ought to be considered:

i.   During their training time, future teachers are not accustomed to using these skills.
ii. There are concerns about the impact that these grades may have on the students' record.
iii. The whole experience is considered trivial.

Relying on all available analysis is necessary to identify each case. This allows to observe certain clues: there are students who always overgrade in the same absolute order of magnitude, regardless of the object and the context of the evaluation, as explained in the third case. There are also cases in which the grade is very different from the teacher's, therefore the object and the context of the evaluation seem to play a role in the evaluation, as per the second case. Therefore, the relative difference of this last assumption remains as an indication to be considered on the ability to use impartial judgment, as per the first case. It goes without saying that all of the above is based on considering the teacher's opinion sufficiently objective which could, by any means, be improved.

Peer-to-peer assessment is essential to support the above-mentioned findings. The general behavior was different from that of the self-assessment. The group of evaluators who over-graded, as compared to the teacher's, was

also the largest. Nonetheless, they seemed to fall more clearly within the second than the third category, given the similar slope of the regression curve and the diagonal of ideal correspondence with the teacher, although what gives its robustness is a smaller percentage that falls within the confidence interval. On the other hand, the undergrading shows a more constant horizontal distribution, which can be associated with the third case, mirroring it in terms of self-evaluation.

One striking aspect in the group assessment is that it does not correspond to that of its members. Thus, in most cases the grades received by the members were not usually similar to the individual group grade; this leads us to think that the students' attitude towards this type of assessment was more oriented towards the third type of those described in the self-evaluation. The typical example here was group J, which received one-third of the lowest scores and held the worst absolute score with just 76 points. However, the average of its members is 88.42, with ID53 as the worst graded (87.26) and ID6 as the best (91.18). Clearly, there is a significant change in attitude when it comes to valuing a group rather than the individuals that make it up as the group tends to receive less favorable grades.

The idea of clarifying or qualitatively explaining the grades did not find much support. And even though the analysis of the few available rated grades made it possible to identify some relatively interesting behaviors, ultimately, this was not a relevant aspect of the 360-degree assessment under study. Therefore, what is relevant in our opinion is to consider why this happened. The most plausible explanation found lies in the complexity of the tool itself, as it had to be evaluated on dozens of occasions and by considering different assessment parameters. Undoubtedly, this requires more time than that what is available during the presentations.

There is therefore a difficult dilemma: this issue could have been given additional time, which would have meant stopping the rhythm of the presentations of final works in the final phase of the course; or it could have been done on a less important assignment, for instance in the middle of the course. Regarding the first possibility, there are solid counterarguments about the strain that this would place on the curricular development of the school subject, with the risk of trivializing the contents necessary to make a sound intellectual judgment about the subject itself. And as for the second, the very nature of the 360-degree assessment would not have been easier to carry out, only the amount of group work prior to it; with the added fact that, focusing the interest on an issue that is perceived as minor, compromises the students' interest and motivation.

As for the last aspect under consideration, i.e., performance by gender, although statistically it offers clearly different behaviors, it must be relativized

for two reasons. The first is the sample bias in the gender ratio, which is clearly more female dominated; the second is magnitude of these discrepancies, which are statistically negligible and move in the range of 2 to 3 points, or 2 to 3 tenths on a traditional decimal scale. However, both this aspect and the previous ones, ought to be repeated several times to determine more reliably whether we are dealing with generalizable patterns, and in what context and to what extent, if any.

Finally, we consider that the research question has been affirmatively answered, as long as the two sub-questions have been satisfied. First, the 360-degree assessment made it possible to identify student behaviors according to the scale of analysis, that could be related to their performance in terms of the acquisition of competencies; in this case the focus was on the development and strengthening of the ability to issue reflective and objective critical judgment among peers. And second, group behavior has been observed for both under and overestimated situations, which gives us clear indications of gregarious bias of judgment regarding. As observed with widespread reluctance to explain the scores in detail.

Undoubtedly, the complexity of the tool resulted in some difficulties, such as the teacher's control over its full application. However, it has also made it possible to create a controlled scenario for a professional valuation that would be difficult to carry out in its full scope by using other methodologies. This is of unquestionable added value in curricular areas such as Social Sciences and their didactics, in which the subjective component is more recurrent and evident than in other fields (Natural Sciences, Mathematics), even projecting itself into the political sphere. Therefore, not only does it help to train teachers who will be fairer in their professional opinion, but also to create freer citizens, as it helps them acquire intellectual tools, they can resort to in order to make the decisions they will have to face in their life.

As previously pointed out, new questions are also arising which call for this experience to be improved and repeated in this and other contexts, in order to find patterns that will ultimately help improve the teaching and learning of curricular contents, something essential in the educational process. This is because -according to Aristoteles and Professor Gustavo Bueno- the same as philosophizing means philosophizing against someone, teaching means teaching against ignorance, superstition, and obscurantism.

References

Álvarez-López, Gabriel, and María Matarranz. 2020. "Calidad y evaluación como tendencias globales en política educativa: estudio comparado de agencias

nacionales de evaluación en educación obligatoria en Europa." *Revista Complutense de Educación* 31 (1): 83-93. https://doi.org/10.5209/rced.61865.

Andrade, G. 2019. Standing up for science against postmodernism and relativism. *Philosophia (Philippines)*, 20, 197-211.

Arancibia Herrera, M. 2016. "Uso educativo de las TIC y su relación con las concepciones del profesor de historia sobre aprender y enseñar." *Revista Infancia, Educación y Aprendizaje* 2 (2): 76-93.

Arreola Rico, Roxana Lilian. 2019. "Formación y evaluación docente basada en un perfil por competencias. Una propuesta desde la práctica reflexiva." *Revista Educación* 43 (2): 239-58. https://doi.org/10.15517/revedu.v43i2.30898.

Báez-Rojas, Claudio, Karen Córdova-León, Lincoyán Fernández-Huerta, Ricardo Villagra-Astudillo, Laura Aravena-Canese, Claudio Báez-Rojas, Karen Córdova-León, Lincoyán Fernández-Huerta, Ricardo Villagra-Astudillo, and Laura Aravena-Canese. 2021. "Modelo de retroalimentación mediante evaluación de 360º para la docencia de pregrado en ciencias de la salud." *FEM: Revista de la Fundación Educación Médica* 24 (4): 173-81. https://doi.org/10.33588/fem.244.1133.

Barba Aragón, María Isabel. 2020. "La satisfacción del alumnado con la evaluación 360º." In *CIVINEDU 2020. 4th International Virtual Conference on Educational Research and Innovation*, REDINE, Red de Investigación e Innovación Educativa, 148-50. Madrid: Adaya Press.

Barbera, J., Naibert, N., Komperda, R., and Pentecost, T. C. 2020. Clarity on Cronbach's Alpha Use. *Journal of Chemical Education*, 98(2): 257-258. https://doi.org/10.1021/acs.jchemed.0c00183

Bujang, M. A., Omar, E. D., and Baharum, N. A. 2018. A Review on Sample Size Determination for Cronbach's Alpha Test: A Simple Guide for Researchers. *The Malaysian journal of medical sciences: MJMS*, 25(6): 85–99. https://doi.org/10.21315/mjms2018.25.6.9

Cebrián Robles, D. 2019. Evaluación formativa para los aprendizajes del prácticum con Corubric. In *XV Symposium Internacional sobre el Prácticum y las Prácticas Externas:" Presente y retos de futuro": actas*, Poio (Pontevedra), 10, 11 y 12 de julio de 2019 (23). Asociación para el Desarrollo del Prácticum y de las Prácticas Externas.

Cebrián-de-la-Serna, Manuel, and M.E. Bergman. 2014. "Evaluación formativa con e-rúbrica: aproximación al estado del arte." *REDU : Revista de Docencia Universitaria* 12 (1): 15-22.

Cubillos-Veja, Carla, and Magdalena Ferrán-Aranaz. 2018. "Diseño y Validación de una Rúbrica para Valorar la Resolución de Casos Prácticos Relativos a Derechos Humanos." *Revista Iberoamericana de Evaluación Educativa* 11 (2). https://doi.org/10.15366/riee2018.11.2.002.

Curcu, Antonio. 2008. "Sujeto, subjetividad y formación en educación para pensar en otra visión pedagógica de la evaluación." *Revista de Teoría y Didáctica de las ciencias sociales* 13: 195-216.

Dagal, Asude Balaban, and Rengin Zembat. 2017. "A Developmental Study on Evaluating the Performance of Preschool Education Institution Teachers with 360 Degree Feedback." *Journal of Education and Training Studies* 5 (6): 220-31. https://doi.org/10.11114/jets.v5i6.2365.

Das, K. 2019. "The role and impact of ICT in improving the quality of education: An overview." *International Journal of Innovative Studies in Sociology and Humanities*, 4(6): 97-103.

Elías, Rudy. 2017. "Los programas internacionales de evaluación de logros académicos y su influencia en las políticas educativas en América Latina." *Población y Desarrollo* 45: 74-82.

Emerson, R. W. 2019. "Cronbach's Alpha Explained." *Journal of Visual Impairment and Blindness*, 113(3): 327.

Fernández-Quero, Juan Luis. 2021. "El uso de las TIC como paliativo de las dificultades del aprendizaje en las ciencias sociales." *Digital Education Review*, n.º 39 (julio): 213-37. https://doi.org/10.1344/der.2021.39.213-237.

Ferreiro Concepción, Jasiel Félix, and Carlos Rafael Fernández Medina. 2020. "Una mirada a la evaluación por rúbricas a través de las TIC." *Mendive*. *Revista de Educación* 18 (1): 92-104.

Frías-Navarro, Dolores, and Marcos Pascual-Soler. 2021. "Recomendaciones para elaborar y redactar el informe de investigación." In *Research design, analysis and writing of results*, Frías-Navarro, Dolores y Pascual-Soler, Marcos. Valencia: OSF. https://doi.org/10.17605/osf.io/kngtp.

Furedi, F. 2018. *Qué le está pasando a la Universidad: Un Análisis Sociológico de su Infantilización*. Madrid: Narcea Ediciones.

Galaz, Alberto, and Sergio Toro Arévalo. 2019. "Estrategias identitarias. La subjetividad del profesor ante la política de evaluación." *Andamios* 16 (39): 353-78. https://doi.org/10.29092/uacm.v16i39.687.

García-Valcárcel Muñoz-Repiso, Ana, Azucena Hernández Martín, Marta Martín del Pozo, and Susana Olmos Migueláñez. 2020. "Validación de una rúbrica para la evaluación de trabajos fin de máster." *Profesorado, Revista de Currículum y Formación del Profesorado* 24 (2): 72-96. https://doi.org/10.30827/profesorado.v24i2.15151.

Gatica-Lara, Fiorina, and Teresita del Niño Jesús Uribarren-Berrueta. 2013. "¿Cómo elaborar una rúbrica?" *Investigación en Educación Médica* 2 (5): 61-65. https://doi.org/10.1016/S2007-5057(13)72684-X.

González-Such, José, Jesús Miguel Jornet-Meliá, María Jesús Perales Montolío, Margarita Bakieva-Karimova, Carlos Sancho-Álvarez, Purificación Sánchez-Delgado, and Sonia Ortega Gaite. 2021. "Evaluación de titulaciones universitarias según su aportación a la cohesión social (UNIVECS). Resultados de un análisis de validación cualitativa a través de grupos focales en dos titulaciones de máster de la Universitat de València." *New Trends in Qualitative Research* 7 (julio): 278-95. https://doi.org/10.36367/ntqr.7.2021.278-295.

Grande de Prado, Mario, Francisco José García Peñalvo, A. Corell, and V. Abella-García. 2021. "Evaluación en Educación Superior durante la pandemia de la COVID-19." *Campus Virtuales* 1 (10): 49-58.

Hussen Maulud, D., Zeebaree, S. R. M., Jacksi, K., Mohammed Sadeeq, M. A., and Hussein Sharif, K. 2021. State of Art for Semantic Analysis of Natural Language Processing. *Qubahan Academic Journal*, 1(2), 21–28. https://doi.org/10.48161/qaj.v1n2a44.

Jogezai, N. A., Baloch, F. A. and Ismail, S. A. M. M. 2020. Hindering and enabling factors towards ICT integration in schools: A developing country perspective. *İlköğretim Online – Elementary Education Online*, 19(3): 1537-1547. https://doi.org/10.17051/ilkonline.2020.733176.

Jornet, José María, María Jesús Perales, and Purificación Sánchez-Delgado. 2011. "El valor social de la educación: entre la subjetividad y la objetividad. Consideraciones teórico-metodológicas para su evaluación." *RIEE* 4: 51-77.

Jornet Meliá, Jesús Miguel, María Jesús Perales Montolío, and José González-Such. 2020. "El concepto de validez de los procesos de evaluación de la docencia: The concept of validity of teaching evaluation processes." *Revista Española de Pedagogía* 78 (276): 233-52.

Kestel, M. and Korkmaz, I. 2019. The impact of Modernism and Postmodernism on Teachers. *Turquoise International Journal of Educational Research and Social Studies*, 1(1): 28-33.

Liu, Lili, Yilin Zhong, Gou Yu, Xinlin Hou, and Jianguang Qi. 2021. "Application of 360-degree assessment in the competency-oriented standardized training of pediatric residents." *Chinese Journal of Medical Education*.

López Aguilar, G. R., M. V. A. Lara Andrade, E. Montiel Piña, M. A. Cruz Gómez, and S. Flores González. 2020. "Evaluación Diagnóstica en la Educación Superior con Prospectiva Educativa para Una Sociedad Incluyente." In *Agenda pública, gobernanza metropolitana y planeación prospectiva para un desarrollo sostenible, incluyente y solidario*, Vázquez Guzmán, O. y Carrillo Huerta, M. M., 90-100. Puebla (México): Benemérita Universidad Autónoma de Puebla. https://bit.ly/3iwVrw8.

López-de-Arana Prado, Elena, Pilar Aramburuzabala Higuera, and Héctor Opazo Carvajal. 2019. "Diseño y validación de un cuestionario para la autoevaluación de experiencias de aprendizaje-servicio universitario." *Educación XX1* 23 (1). https://doi.org/10.5944/educxx1.23834.

Mano, R. 2021. The institutionalization of ICT and civic participation: Evidence from eight European nations. *Technology in Society*, 64, 101518. https://doi.org/10.1016/j.techsoc.2020.101518

Martínez Romera, Daniel. 2017. "La evaluación de diseños de intervención para el Prácticum como instrumento de formación e investigación en el Máster de Profesorado de Ciencias Sociales." *Revista Practicum* 2 (diciembre): 32-49. https://doi.org/10.24310/RevPracticumrep.v2i2.9857.

———. 2019. "Recursos para innovar en geografía e historia: un camino hacia las TAC." In *Tecnologías para la formación de profesionales en educación*, Alías

Garcia, Antonio, Cebrián Robles, Daniel, Ruiz Rey, Francisco José y Caraballo Vidal, Israel. Dykinson. https://doi.org/10.2307/j.ctv105bcxd.18.

Martínez Romera, Daniel, Daniel Cebrián Robles, and Rafael Pérez Galán. 2020. "Practical Training of Secondary School Teachers in Spain: Tutoring and Assessment Using ICT." *Turkish Online Journal of Distance Education* 21 (abril): 153-66. https://doi.org/10.17718/tojde.728148.

Meghdad, Rahati, Rohollahi Nayereh, Sakeni Zahra, Zahed Houriye, and Nanakar Reza. 2020. "Assessment of the performance of nurses based on the 360-degree model and fuzzy multi-criteria decision-making method (FMCDM) and selecting qualified nurses." *Heliyon* 6 (1): e03257. https://doi.org/10.1016/j.heliyon.2020.e03257.

Milliken, J. 2004. Thematic reflections on higher education. *Higher Education in Europe*, 29(1): 9-18. https://doi.org/10.1080/03797720410001673265.

Oliveras Boté, Isabel. 2019. *Evaluación e incorporación del riesgo de sesgo de estudios no experimentales en revisiones sistemáticas y metaanálisis*. Universidad Autónoma de Barcelona. https://ddd.uab.cat/record/207913.

Ortega-Quevedo, Vanessa, Cristina Gil-Puente, Cristina Vallés-Rapp, and María Antonia López-Luengo. 2020. "Diseño y validación de instrumentos de evaluación de pensamiento crítico en Educación Primaria." *Tecné, Episteme y Didaxis: TED* 48 (diciembre): 91-110. https://doi.org/10.17227/ted.num48-12383.

Ortiz-Revilla, Jairo, Ileana María Greca, and Agustín Adúriz-Bravo. 2021. "Conceptualización de las competencias: Revisión sistemática de su investigación en Educación Primaria." *Profesorado, Revista de Currículum y Formación del Profesorado* 25 (1): 223-50. https://doi.org/10.30827/profesorado.v25i1.8304.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825-30.

Perales-Montolío, M.J., J.M. Jornet-Meliá, and J. González-Such. 2014. Tendencias en las políticas de formación y evaluación del profesorado en la educación superior en España. *Revista Iberoamericana de Evaluación Educativa*, 7(2e): 53-64. https://goo.gl/qUxOKT.

Pozuelos Estrada, Francisco José, Francisco Javier García Prieto, and Sara Conde Vélez. 2020. "Evaluar prácticas innovadoras en la enseñanza universitaria. Validación de instrumento." *Educación XX1* 24 (1). https://doi.org/10.5944/educxx1.26300.

Rodríguez-Gómez, David, Carme Armengol, and Julio Meneses. 2017. "La adquisición de las competencias profesionales a través de las prácticas curriculares de la formación inicial de maestros." *Revista de Educación* 376: 229-51. https://doi.org/10.4438/1988-592X-RE-2017-376-350.

Sarceda-Gorgoso, M. Carmen, and María Luisa Rodicio-García. 2018. "Escenarios formativos y competencias profesionales en la formación inicial del profesorado." *Revista Complutense de Educación* 29 (1): 147-63. https://doi.org/10.5209/RCED.52160.

Searle, J. R. 1995. *Posmodernism and the Western Rationalist Tradition*. Routledge.

Shafait, Z., Khan, M.A., Sahibzada, U.F., Dacko-Pikiewicz, Z., Popp, J. (2021). An assessment of students' emotional intelligence, learning outcomes, and academic efficacy: A correlational study in higher education. *PLoS ONE* 16(8): e0255428. https://doi.org/10.1371/journal.pone.0255428

Taber, Keith S. 2018. "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education." *Research in Science Education* 48 (6): 1273-96. https://doi.org/10.1007/s11165-016-9602-2.

Tejada-Fernández, J., J. Serrano-Angulo, C. Ruiz-Bueno, and D. Cebrián-Robles. 2015. "El proceso de construcción y validación de los instrumentos de recogida de información sobre el practicum y su evaluación a través de herramientas tecnológicas." In *Building and Validation of Instruments to Collect Information on the Practicum and their Evaluation Based on ICT Tools*, editado por M. Rapos-Rivas, Muñoz-Carril, M.E.Martínez Figueira Zabalza Cerdeiriña, A., y Pérez Abellas, 261-72. Poio, España: Andavira Editora.

Urquidi Martin, Ana Cristina, María Sol Calabor Prieto, and Carmen Tamarit Aznar. 2019. "Entornos virtuales de aprendizaje: modelo ampliado de aceptación de la tecnología." *Revista Electrónica de Investigación Educativa* 21 (junio): 1-12. https://doi.org/10.24320/redie.2019.21.e22.1866.

US Office of Personnel Management. 1997. *360-Degree assessment: An overview*. United States Office of Personnel Management. https://bit.ly/3RbsR2j.

Usart Rodríguez, Mireia, José Luis Lázaro Cantabrana, and Mercè Gisbert Cervera. 2020. "Validación de una herramienta para autoevaluar la competencia digital docente." *Educación XX1* 24 (1). https://doi.org/10.5944/educxx1.27080.

Vendrell i Morancho, Mireia, and Jesús Miguel Rodríguez Mantilla. 2020. "Pensamiento Crítico: conceptualización y relevancia en el seno de la educación superior." *Revista de la educación superior* 49 (194): 9-25. https://doi.org/10.36857/resu.2020.194.1121.

Zhoc, K.C.H., King, R.B., Chung, T.S.H. and Junjun, C. 2020. Emotionally intelligent students are more engaged and successful: examining the role of emotional intelligence in higher education. *European Journal of Psychology Education* 35: 839–863. https://doi.org/10.1007/s10212-019-00458-0.

Zweekhorst, M.B.M. and Maas, J. 2015. ICT in higher education: students perceive increased engagement. *Journal of Applied Research in Higher Education*, 7(1): 2-18. https://doi.org/10.1108/JARHE-02-2014-0022.

About the authors

DR. DANIEL DAVID MARTINEZ-ROMERA (corresponding author, ddmartinez@uma.es) is Lecturer in Teaching of Social Sciences in the Faculty of Education at the University of Málaga, accredited as an Associate Professor. PhD in Education from the University of Malaga, PhD in Geography from the University of Granada and international postgraduate in Geographic Information Systems from UNIGIS International University. From 2005 to 2010, he was a trainee

researcher and a PhD researcher at the Andalusian Institute of Statistics and Cartography, within an agreement with the University of Granada. Over the last twelve years, his main lines of research are: Geography, Social Sciences and Teaching; Development and Innovation of ICT Applied to Teaching. An example of his work is the Diagrom teaching and analysis assistants (on ombrothermic diagrams) and Piradem assistants (on population pyramids); both available on his personal website. He also conducts quantitative and qualitative analyses of educational contexts and assessment processes supported by ICT.

MTR. SARA CORTÉS-DUMONT (scortes@ujaen.es) is Postgraduate Teaching Assistant in Geography in the Faculty of Humanities at he University of Jaén, Máster en Sistemas de Información Geográfica por la Universidad Internacional en Sistemas de Información Geográfica (UNIGIS). From 2003 to 2011 she was training intern and research fellow at the Andalusian Institute of Statistics and Cartography. She worked as GIS-specialist at University of Córdoba and freelancer until 2019. Her areas of interest cover Human Geography and Teaching innovation, focusing in knowledge transfer to society. She develops this interest as researcher through recent I+D+I National Projects: "Avanzando en la modelización: Fuentes catastrales y paracatastrales en el Antiguo Régimen. Territorio, población, recursos, funciones" (PID2019-106735GB-C22), "Modelización de patrones para la caracterización de la Córdoba eclesiástica del siglo XVIII según el catastro de Ensenada y otras fuentes Geohistóricas" (CSO2015-68441-C2-2-P) and "Dinámicas funcionales y ordenación de los espacios del "Sistema del Patrimonio Territorial andaluz: Análisis en Andalucía Occidental" (CSO2010-19278).