Assessment professional development courses for university teachers: A nationwide analysis exploring length, evaluation and content knowledge.

Javier Fernández Ruiz[1] & Ernesto Panadero[2, 3]

Author Note

[1] Departamento de Psicología Evolutiva y de la Educación, Facultad de Psicología, Universidad Autónoma de Madrid, Spain.

[2] Facultad de Educación y Deportes. Universidad de Deusto, Spain.

[3] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

Correspondence concerning this manuscript should be addressed to: Javier Fernández Ruiz. Aula PDIF, Facultad de Psicología. Universidad Autónoma de Madrid, Cantoblanco. 28049. Spain. E-mail: javier@fernandezruiz.com

The authors declare not having any conflict of interest regarding this manuscript.

Abstract

The design and implementation of assessment is one of the main challenges for university teachers. In previous studies, university teachers have claimed needing a better professional development regarding assessment. With the aim of analyse how university teachers are educated, this study presents the first nationwide analysis of assessment professional development courses. Every professional development course from Spanish public universities (N = 1627) was screened, and data from all available assessment professional development courses (N = 82) was collected and analysed. These courses were compared in terms of total length, evaluation of teachers' learning, and content knowledge covered. Regarding total length, while some universities use a one-session approach, others prefer longer and more intensive courses. Regarding the evaluation of the courses results, most universities do not evaluate teachers' learning and the ones which do it tend to use passive methods such as attendance to the courses. Regarding content knowledge covered, important topics such as self- and peer assessment or feedback are vastly underrepresented. The characteristics of the courses analysed are compared with the literature recommendations concluding that there is a significant room for improvement.

Keywords: assessment literacy; teacher education; assessment

Assessment professional development courses for university teachers: A nationwide analysis exploring length, evaluation and content knowledge.

Assessment practices can have beneficial effects on students' learning, motivation or self-regulation, among many other variables (Andrade and Heritage, 2017; McMillan and Moore, 2020). In many countries, university teachers are responsible for designing and implementing a particular set of assessment practices, so it is important to understand the factors that influence their decisions (Bearman et al., 2017; Adachi, Tai, and Dawson, 2018). Previous research has shown that university teachers struggle when designing and implementing assessment and claim to lack the knowledge to make meaningful innovations (Fernández Ruiz, Panadero, García Pérez, and Pinedo, 2021). To help them to carry out better-informed decisions, one key factor to address is the type of education that university teachers receive in assessment.

For that reason, teacher professional development has been a frequently discussed topic in European universities. Some countries have decided to offer compulsory education (Gibbs and Coffey, 2004) and require their university teachers to undertake professional development courses during their career. However, this is not the case in other countries, such as Spain, where universities do not oblige their teachers to complete structured professional development courses. Nevertheless, it is a common practice for universities to offer elective courses covering a wide range of topics related to research, teaching and management, within a programme teachers can attend voluntarily to earn teaching accreditations which are positive for their CVs.

Although debate remains about the benefits of compulsory teacher education (Trowler and Bamber, 2005), it is essential to ensure that such education, whether compulsory or not, is of the highest quality possible to promote the ability among university teachers to make informed decisions about their practices. To evaluate the

quality of teacher education in Spanish universities, and considering the lack of a national standardised design, it would be necessary to collect and analyse the courses offered by each university which, to our knowledge, has not yet been done. Our study analysed the characteristics of the assessment professional development courses (from now on APDC) offered to teachers in all Spanish public universities.

How to Make Teacher Professional Development in Higher Education Effective

Several authors have explored the characteristics of effective professional development courses for university teachers and how to achieve a real change in their knowledge, attitudes and practices regarding different aspects of their pedagogical practice (e.g. Steinert et al., 2006, 2016; Postareff, Lindblom-Ylänne and Nevgi, 2007; Stes et al., 2010). It should be noted that whether or not these courses can be effective has been a matter of debate among researchers, with studies supporting both perspectives. For example, Gibbs and Coffey (2004) found that long-exposure professional development courses (4 to 18 months) had positive effects on university teachers, as they developed a greater student-centred approach. On the contrary, Norton et al. (2005) did not find any significant results when comparing educated and uneducated university teachers in terms of beliefs and intentions about teaching. They argued that the lack of differences between educated and uneducated university teachers was because not every professional development course is effective in producing real changes in teacher beliefs: only training which meets certain requirements is effective.

In recent years, several reviews have explored these requirements. Two lines of work are of special importance for our study, those by Steinert et al. (2006, 2016) and Stes et al. (2010). First, Steinert et al. (2006) reviewed 53 papers, analysing the effects of professional development courses among medical education teachers. Those effects were categorised in four levels of educational outcomes and were grouped by type of

intervention. Despite its focus on medical education, their proposal has been used in studies oriented to a general population, such as Stes et al. (2010). Adapting the framework proposed by Steinert et al, these authors carried out a review on professional development courses in higher education. In their review, they looked for effects of professional development courses on teachers' learning and behaviour, institutional impact and students' perceptions, study approaches and learning outcomes. These effects were also compared in terms of the length, instructional method and target group of the courses.

Both reviews have listed key features to design effective professional development courses. There are three features which we specifically highlight. The first is the length of the courses, considering the benefits of extended courses over shorter interventions. Steinert et al. (2006) found that longer courses, such as seminar series, with several sessions spaced over time, tended to produce outcomes not found in shorter interventions. Some of these outcomes would be related to the creation of networks and cooperative interactions among teachers (Rawer et al., 1997) or a higher involvement in educational activities. In their follow-up review (Steinert et al., 2016), the benefits of longer courses are also reported. In a similar vein, Stes et al. (2010) concluded that prolonged education was more effective than one-off events. This difference is especially observed in the behaviour of the teachers, although these authors claimed the need to take these results with caution due to the small sample size.

The second key feature is the importance of supervised practice and feedback. In the review by Steinert et al. (2006), several authors highlighted the importance of applying what has been learned, practising skills while also receiving feedback on those skills (Coles and Tomlison, 1994; Hewson, 2000). In their follow-up review (Steinert et al., 2016), the importance of the opportunities for practice and application while

receiving feedback was also highlighted. The opportunity for supervised practice is especially relevant, considering the difficulties reported by university teachers when implementing changes in their assessment methods (Fernández Ruiz, Panadero, García Pérez, and Pinedo, 2021). To be precise, these difficulties can be explained by various contextual and personal pressures (Yorke, 2003), and could be alleviated to a great extent through having a controlled space in which to try assessment innovations.

The third key feature mentioned in previous reviews (Steinert et al., 2006, 2016) is the importance of evidence-informed design, which must integrate a theoretical or conceptual framework to guide university teachers' learning. In the field of educational assessment in higher education, several frameworks have been proposed, which discuss the importance of certain strategies to make assessment more sustainable and beneficial for student learning (Boud, 2000; Boud and Soler, 2016).

How to design assessment in higher education has been a widely discussed topic in the scientific literature for decades. In the next section, we summarize some of the strategies or procedures which literature mentions as necessary to address to achieve a high-quality assessment.

Assessment in Higher Education

Authors such as Boud (2000) or Sadler (2005) have contributed to highlight the importance of several topics regarding assessment in higher education. Coherently with their weight in the recent scientific literature, a strong presence of these topics in the APDC would be expected. The topics are, as follows:

1. *Self-assessment*: According to Boud (2000), self-assessment is one of the requirements to achieve sustainable assessment in higher education. Authors such as Bourke (2014, 2018) have argued that self-assessment is an important pedagogical approach to support university students in understanding their own

learning. Additionally, several empirical and review works have shown its relationship with academic achievement (respectively, Topping, 2003; Jay and Owen, 2016).

2. *Peer assessment*: The use of peers in assessment is also proposed by Boud (2000) as a requirement for making assessment sustainable. Peer assessment in higher education has been a relevant area of research in the last years, after the seminal review by Topping (1998). Now, it is considered not only a way of promoting student learning (Double, McGrane, and Hopfenbeck, 2020) but also a great opportunity for reducing teachers' workload (Higgins, Grant, and Thompson, 2010).

3. *Feedback*: This is considered a key promoter of student learning (Hattie and Timperley, 2007; Wisniewski, Zierer, and Hattie, 2020); however, its implementation in higher education is a consistent critique received by various educational systems (Boud and Molloy, 2013).

4. *Grading*: The use of numerical grades for summative purposes is a requirement in many higher education systems. In recent years, the literature has looked towards the use of grades with a more formative flavour (Broadbent, Panadero and Boud, 2017). First, the use of standards-based grading has been recommended (Sadler, 2005), with the aim of making grades informative to students, and the advantages of self-grading have also been discussed (Nieminen, 2020).

5. *Disciplinary differences*: Higher education, unlike other educational stages, is strongly organised around disciplinary traditions. Recent studies have shown that disciplinary differences can determine how assessment is designed, or what

makes feedback productive (Fernández Ruiz, Panadero, and García-Pérez, 2021; Esterhazy, 2018).

All these factors are important areas of discussion in the assessment-related literature. According to the criteria proposed in previous reviews (Steinert et al., 2006; 2016), to design effective APDC, the content covered must be aligned with the current knowledge on educational assessment and not limited to one particular and subjective view of how assessment must be done.

So far, we have explored the previous literature on teacher professional development in higher education, trying to extract key features which are demonstrated to make it more effective. From previous work (Steinert et al., 2006, 2016; Stes et al., 2010), we know three key features which can enhance the effectiveness of professional development courses for university teachers: the presence of extended courses, the opportunity for supervised practice and the inclusion of a solid framework guiding assessment of content knowledge.

Unfortunately, despite the vast knowledge on how to design assessment in higher education, university teachers still do not perform assessment in the most powerful ways, as they often face the challenge of implementation. Studies have shown that university teachers struggle to translate their instructional goals into actual assessment practice (Norton et al., 2010; Fernández Ruiz, Panadero, García Pérez, and Pinedo, 2021). Some authors have argued that the low assessment literacy levels of university teachers are having an impact on assessment in universities (Price et al., 2011), so it is important to ensure that the APDC offered to higher education teachers is effective to overcome these shortcomings. In this study, we analysed the status of the three key features mentioned above in the APDC offered in Spanish universities.

Context and Research Questions

This study was conducted in Spain, where the pre-service education of university teachers is clearly insufficient (Sánchez-Moreno and Mayor-Ruiz, 2006). Many begin their work as teachers with little or no knowledge about assessment. Additionally, some authors reported that APDC are not necessarily based on scientific foundations (Ibarra Sáiz and Rodríguez Gómez, 2010; Palacios and López-Pastor, 2013; Brown, 2015). Generally, universities are responsible for designing APDC, and they also establish what requirements, if any, teachers must meet to pass these courses.

There are studies which show that Spanish university teachers update their knowledge, skills and attitudes less frequently than other European teachers (Quesada-Pallarés, Fernández de Álava and Gairín, 2017). In addition, Spanish university teachers have stated that they have difficulty in designing effective and efficient assessment methods (Fernández Ruiz, Panadero, García Pérez, and Pinedo, 2021). Given this, it is crucial to ensure that university teachers receive high-quality APDC and to identify areas for improvement. To do so, we explored three research questions (RQs):

RQ1. What is the length of assessment professional development courses and assessment professional development programmes?

RQ2. What types of evaluation are used in assessment professional development courses?

RQ3. What type of content knowledge is included in assessment professional development courses?

Methods

Sample and Procedure

We screened every professional development course at all 50 Spanish public universities, for a total of 1,627 courses. We considered courses as those sessions or sets of sessions with a specific instructor, chronology and places offered. Such courses

covered a wide range of topics from all areas of a university teacher's professional

activities (teaching, research, and management).

After reading all of the course titles to detect common terminologies, those that

contained the words "assessment" OR "feedback" OR "examination" OR "test" were

selected for further analysis. The decision to limit our analysis to courses which only

topic was assessment was made due to the aim of our study. For example, in courses

about teaching methods in which one section covered assessment, the documents would

not detail how much time was dedicated to each topic, nor what instructional or

evaluation methods were used. This selection yielded 110 courses from 31 universities.

Unfortunately, as not every course had complete information available, the final sample

was 82 courses from 25 universities. The distribution of universities across Spanish

regions is presented in figure 1.

Figure 1

University distribution across Spanish regions
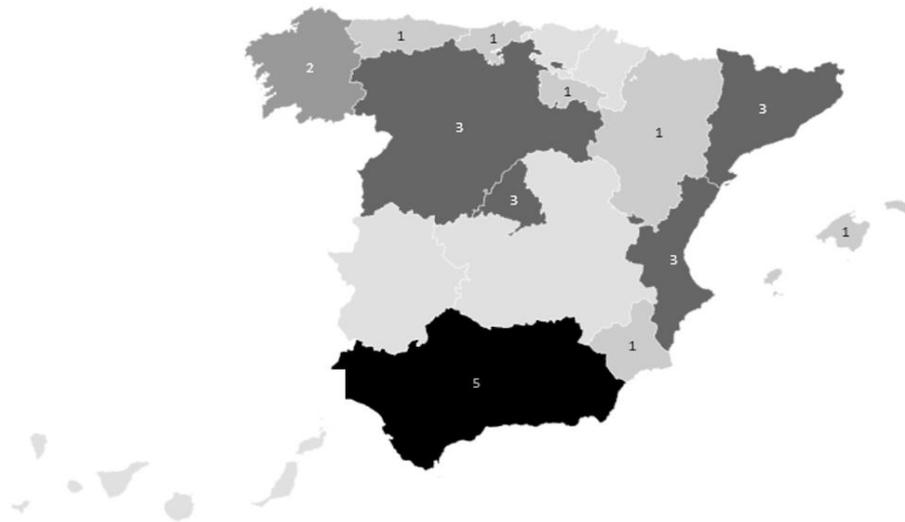
Empirical Planning and Data Extraction

Syllabi from the 82 courses were downloaded and analysed. A syllabus is an explanatory document about a course that is usually available online in most universities. It includes data about the characteristics, contents, instructors and chronology of the course.

As shown in figure 2, we extracted different data for each research question. To answer RQ1, we extracted data regarding course length in hours. To answer RQ2, we extracted data about the evaluation methods used in the course for evaluating teachers' learning. To answer RQ3, we performed a thematic analysis of the learning goals or content of the courses.

Figure 2

Example of syllabus and information extracted
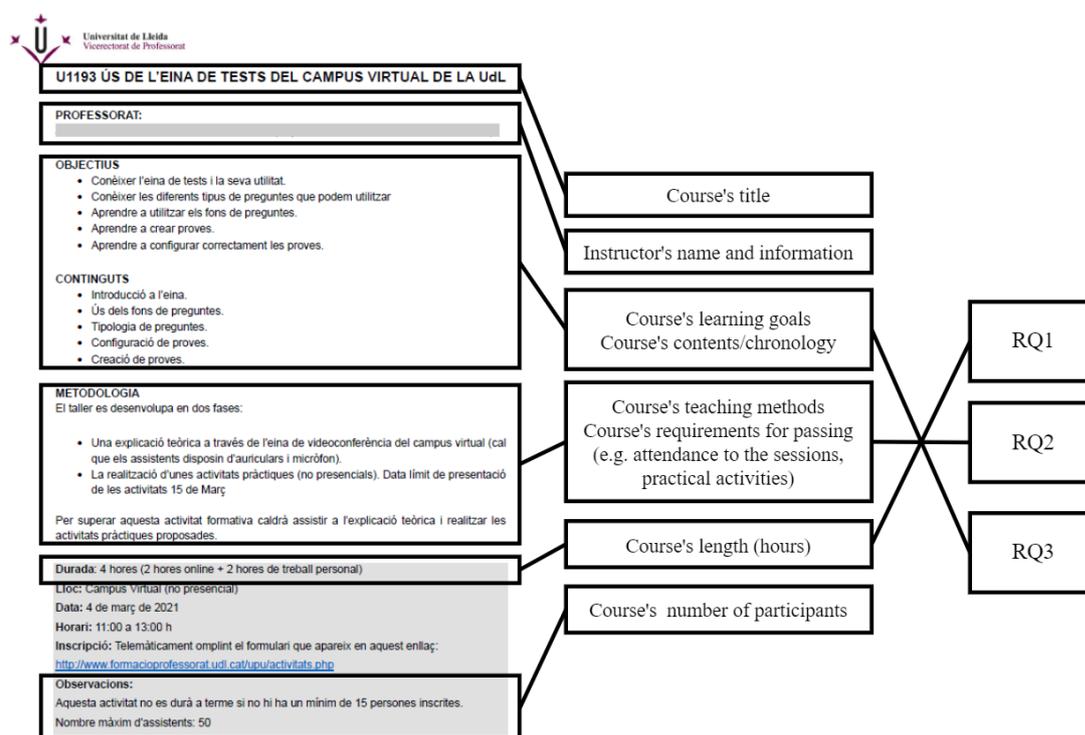
Note: The syllabus shown here belongs to a teacher professional development course from the Universitat de Lleida. Permission for use was granted by the corresponding instructors.

We selected this specific set of data for two reasons. The first is the importance of these variables in previous research (Steinert et al., 2006; Stes et al., 2010), which identified them as key features in successful professional development courses. The

second is that they can be easily extracted from course syllabi. Our research questions proposed three key concepts to be analysed: Length of the courses, evaluation used, and content knowledge covered. The exact information extracted from the syllabi and how it was interpreted to answer each question are discussed below.

The structure of the syllabi across different universities was consistent in general terms. As shown in figure 2, each syllabus includes information about the course title and instructor, learning goals and content, teaching and evaluation methods and logistical information such as length and place of the course. Data about the length of the courses were collected for RQ1. These data were presented in number of hours and analysed descriptively. For RQ2, we explored the type of evaluation used for measuring university teachers' learning during the course. For RQ3, the learning goals of each course were considered. The syllabus for each course included a set of learning goals. Each of these learning goals was considered independently, for a total of 264 learning goals. The courses content was analysed in those cases where the learning goals were not available.

Data Analysis

For RQ1, the hours for each course were extracted directly from the syllabus. These hours were analysed at a descriptive level using the software SPSS 24, including maximum, minimum, mean, median and standard deviation. Additionally, the number of courses offered by each university was also analysed.

For RQ2, the evaluation methods of the courses were distributed in five categories: examinations, assignments, classroom simulation practices, attendance, and teaching innovation projects. One course may have more than one evaluation method for their teachers, so all categories were coded independently and dichotomously depending on their presence or absence in the course syllabus.

For RQ3, we followed an open coding procedure. We screened every course's learning outcome and performed thematic analysis. From this first round, 25 themes were categorized. In the second round of thematic analysis, all 25 themes were grouped into six encompassing areas: (a) assessment methods and tools, (b) assessment of competences, (c) assessment theory, (d) design and implementation, (e) assessment in specific contexts and (f) online assessment. A total of five learning outcomes did not fit any of the previous categories and were labelled as "other".

All RQs were answered in two levels of analysis: "results by course" included data at the course level, and "results by university", included data at the university level. We refer to "courses" when we describe separate courses, and to "programmes" when we describe the set of courses offered by one university. University names are presented in acronyms to improve readability. A full list of the acronyms used is presented in table 1.

Table 1

List of acronyms

Acronym	University
UMA	Universidad de Málaga
UAM	Universidad Autónoma de Madrid
ULL	Universidad de la Laguna
UNIA	Universidad Internacional de Andalucía
UCM	Universidad Complutense de Madrid
UA	Universidad de Alicante
UGR	Universidad de Granada
UV	Universidad de Valencia
UM	Universidad de Murcia
UPO	Universidad Pablo de Olavide
UDC	Universidad de La Coruña
UBU	Universidad de Burgos
UIB	Universidad de las Islas Baleares
UNIOVI	Universidad de Oviedo
UAL	Universidad de Almería
UMH	Universidad Miguel Hernández
UDL	Universitat de Lleida
UNICAN	Universidad de Cantabria
UNIRIOJA	Universidad de La Rioja
USC	Universidad de Santiago de Compostela
USAL	Universidad de Salamanca

|	
UNIZAR	Universidad de Zaragoza
UVA	Universidad de Valladolid
UPM	Universidad Politécnica de Madrid
UPF	Universidad Pompeu Fabra

Results

We will next describe the findings of our research questions. However, it is important to address first the fact that 19 of the 50 analysed universities, do not offer any specific APDC to their teachers. This number represents more than a third of all the Spanish public universities. This is a concerning finding, which is important to highlight before addressing each specific research question.

RQ1. What is the length of assessment professional development courses (APDC) and assessment professional development programmes?

Results by course. There was great variability in the length among the courses analysed. The average was 8.89 hours (SD = 9.15), range between 45 minutes and 40 hours, with a median duration of 6 hours. However, as the standard deviation shows, there is great variability per course in terms of length, as illustrated in figure 3.

Figure 3

Histogram of length of the courses
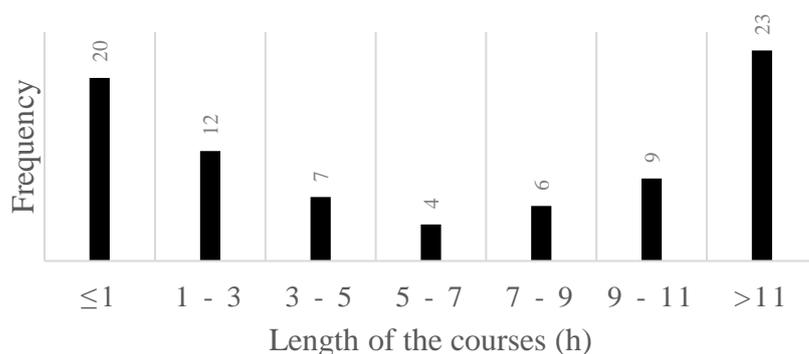
Results by university. As shown in Figure 4, the vast majority of universities are located between the range of one to six APDC, with a notable outlier at UM. The total length of the programmes -calculated summing all individual courses length for each university-

also tends to be less than 30 hours. However, there were universities that did not follow

this trend. On the one hand, universities such as UM seemed to have opted for a

massive one-session format, while also offering many places on their courses, which

nevertheless tend to be short. In the case of this university, we found 23 different

APDC, with a mean duration of 1.13 hours. On the other hand, universities like UMA,

UAM or UGR offered fewer but much longer courses.

Figure 4

Number of courses and total length of training programmes in each university

Number of courses vs total length
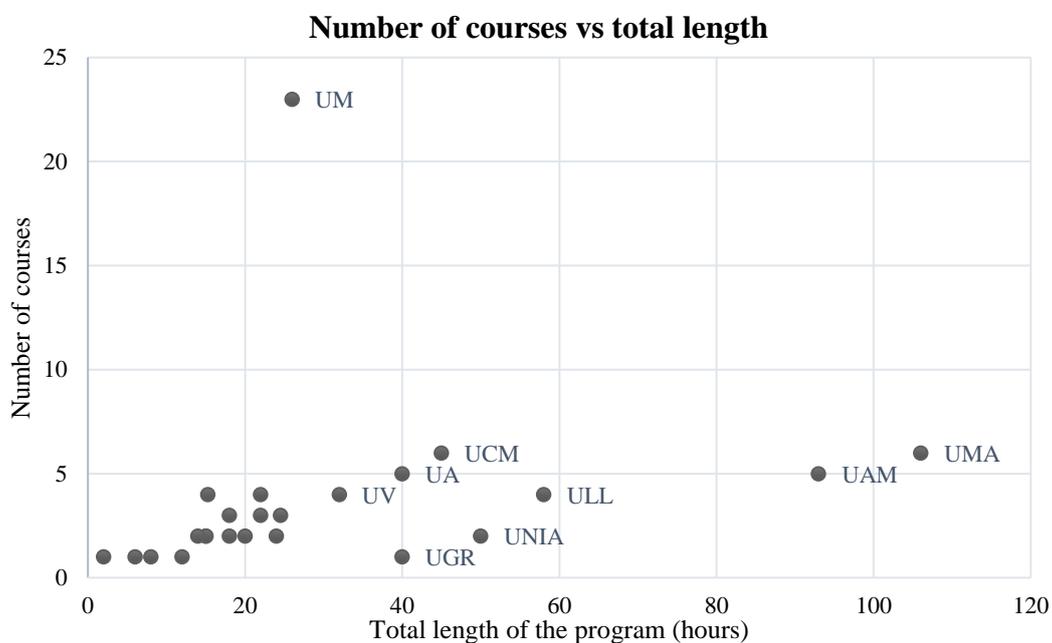
Note: Only universities with complete data available have been included in the figure.

RQ2. What types of evaluation are used in assessment professional development

courses (APDC)?

> ***Results by course.*** The analysis revealed that 47 of the courses included some

form of evaluation of the participant teachers' learning, which was usually associated

with obtaining a certificate. It should be noted that the same course could use several

evaluation practices. For 24 of the courses, attendance was used as an evaluation

method, 17 courses required practical work (e.g., rubric design, learning management

system [LMS] workshops), 14 courses required assignments, 6 courses included

teaching innovation projects and 5 courses had a multiple-choice examination at the

end.

   Results by university. When running the comparison by university, we found

that UMA is the university with more evaluation practices (10), mostly related to

supervised assignments (5) or course attendance (4). In a lower step are UAM and

UCM, with 9 and 8 evaluation practices respectively. In the case of UAM, evaluation

methods are distributed between attendance (4), assignments (3) and practices (2).

UCM, on the contrary, evaluated mostly through teachers' course attendance (6) except

for 2 practical tasks.

   On the contrary, we found five universities which did not evaluate teachers'

learning in their training programmes in any way: UNIA, UM, UDC, UNIRIOJA and

UPM. In most cases, such universities offered just a few APDC, which would explain

the low number of evaluation methods involved. However, in the case of UM, 23

different APDC are reported, with zero assessment planned to evaluate the extent to

which the participant teachers have mastered the course content.

RQ3. What type of content knowledge is included in assessment professional

development courses (APDC)?

   Results by course. As shown in figure 5, a total of 25 themes were identified for

the courses analysed, which were organised into six general areas.

Figure 5

Content knowledge covered in assessment-related courses
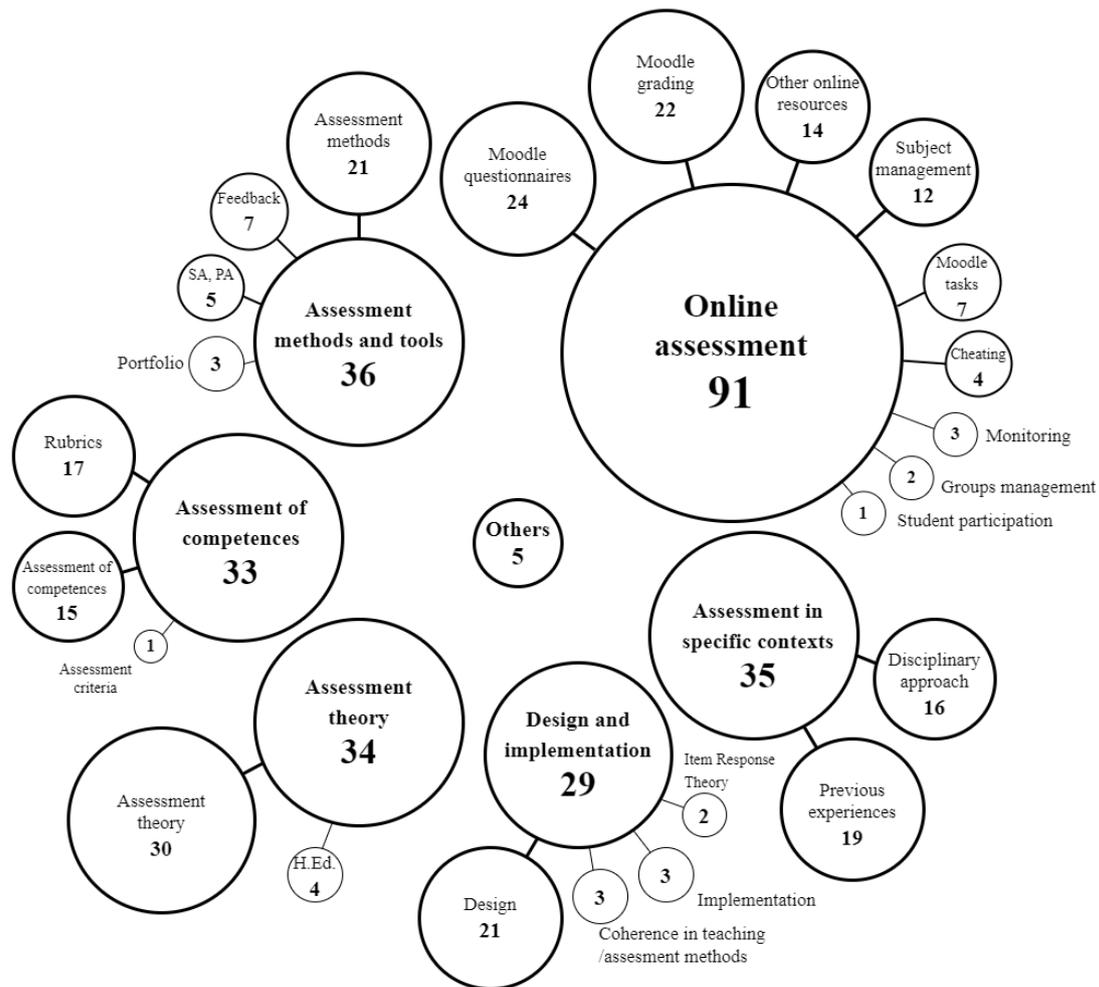
The first main area is "assessment methods and tools", which covers different assessment tasks such as exams, assignments, and portfolios. This area also includes feedback to students, which appeared in seven of the courses. Finally, this area includes education on aspects related to self-assessment and peer assessment.

The second area is "assessment of competences", which is focused on defining the concept of competence and on how to use rubrics to evaluate whether students have reached the required level of competence for a unit. One additional course offered instruction about designing assessment criteria and linking them to specific competences.

The third area is "assessment theory", which focuses on the general literature about assessment, the definitions of and differences between formative and summative

assessment and the particular nature of assessment in higher education and how it differs from assessment in other educational stages.

The fourth area is "design and implementation", which includes more practical advice for teachers in carrying out their preferred assessment methods. Several courses addressed how to design an assessment methodology. Other subtopics in this area are related to the coherence between teaching and assessment methods. Other less common sections explained how to achieve an effective implementation of the designed assessment method, or how to use Item Response Theory to design multiple-choice examinations.

The fifth area is "assessment in specific contexts", and here we found two large blocks. The first one contains APDC with a disciplinary perspective (e.g., focused on assessment in health sciences or in STEM disciplines). The second block is related to the experiences of colleagues in the profession and examples of good practice.

The last area is "online assessment"; within this category, several of the uses of a learning management system (LMS) are detailed, such as questionnaires, grades, tasks and subject management. There were also other courses aimed at preventing contract cheating in online teaching or dedicated to presenting online resources to teachers.

Results by university. Regarding distribution by university, as can be seen in Figure 3, several universities, such as USC, USAL and UNIZAR, focused their programmes exclusively on online assessment. At another level, we found universities, such as UDC or UVA, that focused only on two or three of the areas. Lastly, universities such as UA or UAM offered diverse programmes, covering practically all the areas found. None of the universities offers APDC covering the six areas. The complexity of the assessment literacy content knowledge is well known, so it would be necessary for universities to offer a wide range of topics in their programmes to train

their teachers properly in all aspects of assessment, but this only occurs in a minority of

universities.

Figure 6

Content knowledge distribution in university programmes
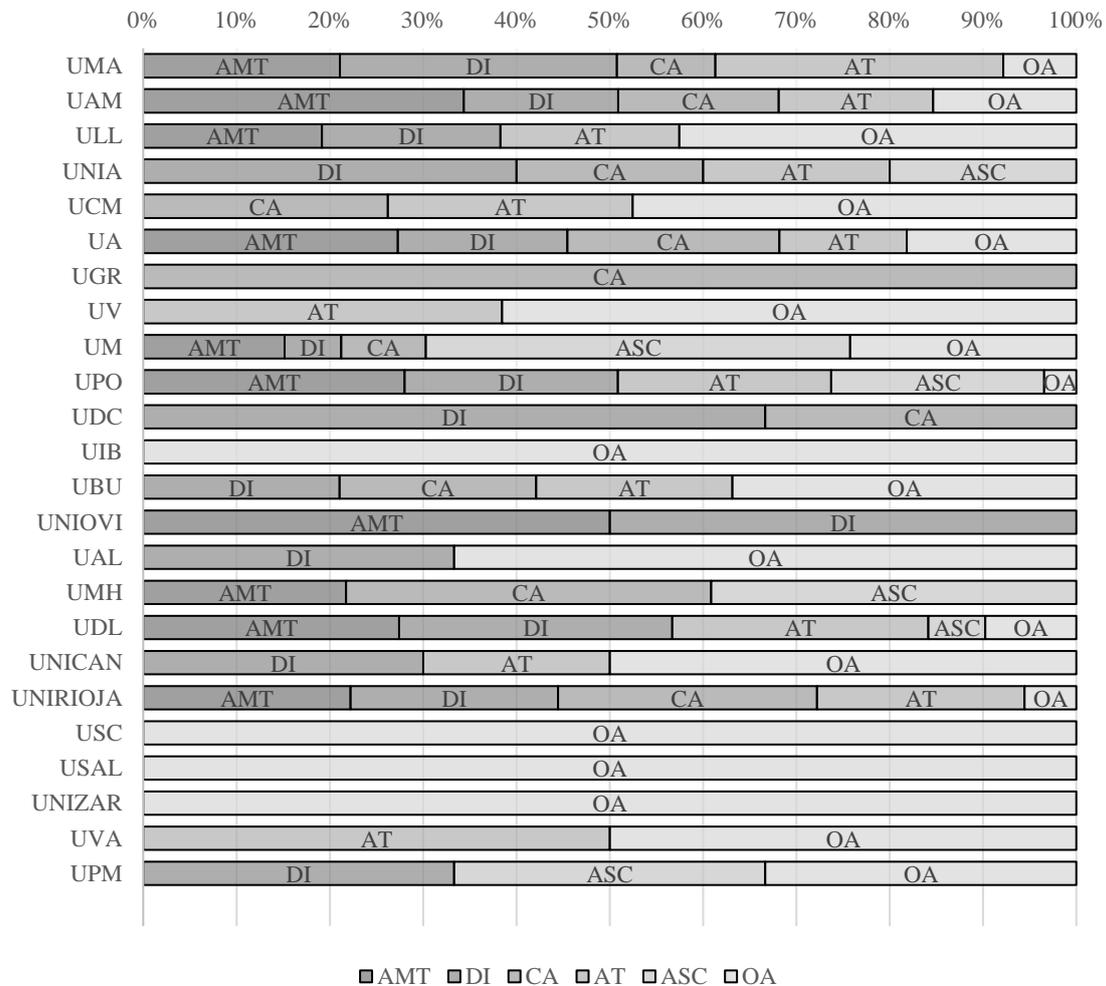
Note: Only universities with complete data available have been included in the figure. Percentage is measured by the total of courses' themes. AMT = Assessment methods and tools; DI = Design and implementation; CA = Assessment of competences; AT = Assessment theory; ASC = Assessment in specific contexts; OA = Online assessment.

As a summary of all our findings, we observed different trends among the

universities for the three areas analysed. Regarding the length of the APDC, some

universities chose to hold several one-session courses, while others preferred small-

group intensive courses. Regarding type of evaluation, most universities offered some

kind of task for the evaluation of participant teachers' learning; however, they are

mostly related to course attendance or the retention of theoretical content. Lastly,

regarding the content knowledge covered, we found a variety of areas with strikingly different distributions. Online assessment was the most discussed topic, while several universities focused all their training programmes on how to implement it in their LMSs. By contrast, relevant areas such as feedback or students' role in assessment were vastly underrepresented.

Discussion

The present study analysed the characteristics of APDC related to assessment offered by Spanish universities. Compared to the recommendations found in the literature (Boud, 2000; Steinert et al., 2006, 2016; Stes et al., 2010) and discussed in previous sections, our results showed that the characteristics of such APDC are, in general terms, underwhelming. First, it was found that 19 universities did not offer any specific APDC, which constitute a striking result. Regarding the remaining universities which do offer APDC to their teachers, we discuss the status of each research question separately.

Length of the Assessment Professional Development Courses and Programmes

Our first research question explored the length of the APDC, as this is one of the main variables related to the success of such courses (Gibbs and Coffey, 2004; Steinert et al., 2006; 2016; Stes et al., 2010). Our results showed great variability in both individual courses and university programmes.

Regarding the universities which do offer APDC, we found important differences. For example, UMA offered several courses of more than 20 hours to a limited number of university teachers (up to 40), while other universities like UM, probably in response to the COVID-19 outbreak during 2020, opted to offer a large number of massive webinars.

The use of webinars as an instructional setting for higher education teachers has been recently explored due to the pandemic outbreak in 2020. As some authors have mentioned, university teachers have usually expressed satisfaction with such methods, as they promote different methods of learning and engagement (Polanco-Bueno, 2013). Meta-analytic studies have also shown how webinars and face-to-face classroom teaching are comparable in their effectiveness to promote student learning (Means et al., 2013; Gegenfurtner and Ebner, 2019). However, the ratio of hours and places in most of the courses analysed suggests that their ability to produce personalised learning, and the opportunity of carrying out supervised practice, is compromised.

The benefits of including several instructional methods in the courses have been discussed in the previous literature (Steinert, 2006). While longer courses can offer a wider variety of instructional methods, it is known that university teachers might not have the necessary time to engage in such long intensive courses (Anderson, 2006). In this case, a mixed offer of short webinars and personalised in-depth courses could be a reasonable approach.

However, assessment is a complex skill which can require time to be mastered. It is necessary to provide teachers with enough time to properly engage in the development of their assessment skills. Therefore, a probably more productive approach, would be to organize the APDC at times when university teachers do not have teaching responsibilities, to ensure that they have the time to participate in more demanding courses. These could be the beginning and end of semesters. Another option would be to design APDC to run along the semester in shorter working sessions, to allow teachers to dedicate time to their professional development in the long term, while keeping their workload balanced.

Evaluation of Assessment Professional Development Courses and Programmes

As has previously been discussed by other authors (Steinert et al., 2010, 2016), experiential learning is extremely relevant to the success of teacher professional development courses. University teachers face tensions when implementing their assessment methods (Carless, 2012; Fernández Ruiz, Panadero, García Pérez, and Pinedo, 2021), and the opportunity to deal with those tensions while being supervised by an expert or their peers is one of the major benefits of in-service courses. In the courses analysed, the outlook on this is concerning. In most cases, the only activity which university teachers had to perform to pass the course was just to attend the sessions. Seventeen courses required a practical work, done mostly in simulation settings (e.g. rubric or questionnaire design), and in only six courses did the evaluation include some form of innovation in the teachers' real classrooms.

We next propose three tentative explanations to explore the lack of a stronger evaluation of the results of the courses which, at the end, negatively affects teachers' learning from these courses. First, in Spain APDC are not compulsory to the teachers, so their participation cannot be taken for granted. The personnel responsible for the teachers' professional development programmes try to attract teachers to those courses, so they tend to design these courses to be appealing to the teachers. Considering the high workload of higher education teachers, the courses requirements limit the amount of work required to pass the course, in an attempt to increase teachers' participation which, unfortunately, might be having a negative effect on teachers' learning. For example, just attending to a course does not mean learning has taken place, as we well know from the evidence on students' attendance.

Second, APDC instructors are in many cases experts coming from other universities, which travel to deliver the session/s. This makes difficult to plan long-term interventions, which would be necessary for teachers to carry out supervised

innovations in their classrooms. These would be the ones that would ensure transfer has taken place.

Lastly, long-term evaluation of participant teachers' learning would require a great effort by the teaching innovation units of each university. It is common practice in most Spanish universities to have a group of specific workers in charge of designing and organizing the teachers' professional development programme. However, in some cases these groups are made up of professionals with limited pedagogical education and not prepared to conduct a specific follow-up evaluation of the courses. Additionally, monetary, and logistical resources are often too tight to enable long-term monitoring, which would be necessary to ensure the transfer of the APDC contents to teachers' practices.

Content Knowledge Included in Assessment Professional Development Courses and Programmes

Our third research question explored the content knowledge covered by the APDC, in comparison with the previous literature about assessment in higher education (Boud, 2000; Boud and Soler, 2016; Esterhazy, 2018). Using thematic analysis, six general areas were identified: assessment methods and tools, assessment of competences, assessment theory, design and implementation, assessment in specific contexts and online assessment. On the one hand, some topics were widely represented, such as those related to LMSs. Online assessment was the most common topic in 10 of the 25 universities analysed, with four of them dedicating their training programmes exclusively to it. Undoubtedly, this was a product of the COVID-19 outbreak, which created an urgent need for university teachers to adapt to online alternatives for assessment (Raza et al., 2021). On the other hand, three important absences were found in the training courses explored, which will be discussed next.

First, student involvement in assessment via self- or peer assessment is practically omitted from training courses (discussed only in five courses), although self-assessment is considered a key strategy to achieving sustainable assessment in higher education (Boud and Soler, 2016). Recent research has shown how its use has positive effects on students' self-regulation and self-efficacy (Panadero, Jonsson, and Botella, 2017) and is also related to academic achievement (Brown and Harris, 2013). The benefits of peer assessment in student learning (Dochy, Segers, and Sluijsmans, 1999) and domain-specific skills (Van Zundert, Sluijsmans, and Van Merriënboer, 2010) are also well-known. Therefore, teachers could make good use of knowledge and skills about its implementation if they were educated on it. Unfortunately, this does not seem to be happening in most universities (Panadero et al., 2019; Lipnevich et al., 2021).

Second, the low presence of feedback content (discussed only in seven courses) is striking. Feedback greatly influences student learning (Wisniewski, Zierer, and Hattie, 2019). From the wide empirical research that has been performed, we know that offering high-quality feedback to students is an extremely complex skill (Hattie and Timperley, 2007; Lipnevich and Panadero, 2021), and its processes are usually misunderstood by university teachers (Boud and Molloy, 2013; Evans, 2013). Considering the importance of feedback in educational research and its strong links with academic achievement, its near absence in the training of university teachers is concerning.

Third, most programmes did not offer discipline-specific training on assessment. Disciplines are still a fundamental organising feature of universities, and their assessment practices are usually different on several levels, such as assessment evidences used (Panadero et al., 2019) or the quantity and type of feedback offered (Jessop and Mackelar, 2016). These differences can be explained because university

teachers from different disciplines work with different aims and challenges when they design and implement their assessment methods (Fernández Ruiz, Panadero, and García-Pérez, 2021). Therefore, university teachers might find highly beneficial the opportunity to attend discipline specific APDC, where their needs and challenges are better represented and contextualised.

The absence of such themes among the APDC could help to explain assessment trends among European universities, including poor-quality feedback, little use of self- and peer assessment and traditional assessment methods (Gómez, Sáiz, and Jiménez, 2013; Panadero et al., 2019). It would be important to rethink the content knowledge about assessment offered to university teachers. As Steinert et al. (2006) have argued, the presence of a solid framework to structure the content knowledge of professional development courses is a decisive factor in their success. In the field of educational assessment, we have reviewed several variables documented in the literature (Boud, 2000; Boud and Soler, 2016; Esterhazy et al., 2018) which are, at best, vaguely represented in the courses analysed.

However, it is also important to acknowledge that, even assessment-literate teachers can face difficulties designing and implementing their assessment methods. There are factors such as university regulations, departmental traditions or even national policies which are strong shapers of educational practices (Bearman et al., 2017; Fernández Ruiz, Panadero, and García-Pérez, 2021). Giving all the credit for traditional assessment methods to the lack of proper APDC offered to university teachers would be a simplistic take. Even so, teachers do claim to have a need for better knowledge about assessment. Offering them that opportunity would eventually lead to significant improvements, even if they are done in small contexts.

Limitations

This study presents several limitations. First, the data collection was performed in February 2021, when educational contexts were strongly affected by the COVID-19 pandemic outbreak. This may have affected the validity of our results, especially those regarding the content knowledge covered. The presence of online assessment courses might not have been so strong in pre-pandemic years.

Second, our data come from official documents and not from direct observation of what the instructors did in the courses. Nevertheless, as Panadero et al. (2019) have argued, the syllabus is a contract between the teacher and students, so it should significantly resemble what happens in the classroom.

Third, the size of the universities might be related to the amount of training offered. Unfortunately, the authors have not been able to obtain reliable data about the current size of the universities studied. Therefore, analysis including university size as a moderator variable has not been performed and remains as a future line of research.

Conclusion

This study presents a nationwide analysis of assessment professional development courses and programmes for Spanish university teachers. When contrasted with the previous literature, it was concluded that these programmes have three areas needing improvement. First, their length is limited and not consistent with the importance of assessment as a pedagogical tool. Second, they use superficial methods of evaluating teachers' learning in the courses. Third, their content is mostly oriented to online assessment and LMSs and omits several key topics on assessment in higher education. These flaws can be considered as evidence for understanding the traditional assessment panorama in Spanish universities and the struggles of Spanish university teachers to design and implement assessment innovations. Ultimately, this study shows how assessment professional development in Spanish universities is a good reflection of

assessment practices: both present areas for improvement to rise to the status of current research.

References

Adachi, C., Tai, J. H. M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education*, *43*(2), 294-306. doi:10.1080/02602938.2017.1339775.

Anderson, G. (2006). Carving out time and space in the managerial university, *Journal of Organisational Change Management*, 19(5), 578-592. doi:10.1108/09534810610696698.

Andrade, H. L., & Heritage, M. (2017). *Using formative assessment to enhance learning, achievement, and academic self-regulation*. New York, NY: Routledge.

Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education*, *74*(1), 49-64. doi:10.1007/s10734-016-0027-7.

Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, *22*(2), 151-167. doi:10.1080/713695728.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation In Higher Education*, *38*(6), 698-712. doi:10.1080/02602938.2012.691462.

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, *41*(3), 400-413. doi:10.1080/02602938.2015.1018133.

Bourke, R. (2014). Self-assessment in professional programmes within tertiary institutions. *Teaching in Higher Education*, *19*(8), 908-918. doi:10.1080/13562517.2014.934353.

Bourke, R. (2018). Self-assessment to incite learning in higher education: developing ontological awareness. *Assessment & Evaluation In Higher Education*, *43*(5), 827-839. doi:10.1080/02602938.2017.1411881.

Broadbent, J., Panadero, E., & Boud, D. (2018). Implementing summative assessment with a formative flavour: A case study in a large class. *Assessment & Evaluation*

In Higher Education, 43(2), 307-322.

https://doi.org/10.1080/02602938.2017.1343455

Brown, S. (2015). *Learning, teaching and assessment in higher education: global perspectives*. London: Palgrave-Macmillan.

Brown, G. T., & Harris, L. R. (2013). *Student self-assessment.* In J. H. McMillan (Ed.), SAGE handbook of research on classroom assessment (pp. 367–393). Los Angeles, CA: SAGE.

Carless, D. (2012). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Abingdon: Routledge.

Coles, C. R., & Tomlinson, J. M. (1994). Teaching student-centred educational approaches to general practice teachers. *Medical Education*, *28*(3), 234-238.

Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, *24*(3), 331-350. doi:10.1080/03075079912331379935.

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*(2), 481-509. doi:10.1007/s10648-019-09510-3.

Esterhazy, R. (2018). What matters for productive feedback? Disciplinary practices and their relational dynamics. *Assessment & Evaluation in Higher Education*, *43*(8), 1302-1314. doi:10.1080/02602938.2018.1463353.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, *83*(1), 70-120. doi:10.3102/0034654312474350.

Fernández-Ruiz, J., Panadero, E., & García-Pérez, D. (2021a). Assessment from a disciplinary approach: design and implementation in three undergraduate programmes. *Assessment in Education: Principles, Policy & Practice*, *28*(5-6), 703-723. doi:10.1080/0969594X.2021.1999210.

Fernández Ruiz, J., Panadero, E., García- Pérez, D. & Pinedo, L. (2021b). Assessment design decisions in practice: Profile identification in approaches to assessment design. *Assessment & Evaluation in Higher Education*, 47 (4), 606–621. doi:10.1080/02602938.2021.1937512

Gegenfurtner, A., & Ebner, C. (2019). Webinars in higher education and professional training: A meta-analysis and systematic review of randomized controlled

trials. *Educational Research Review*, *28*, 100293. doi:10.1016/j.edurev.2019.100293.

Gibbs, G., & Coffey, M. (2004). The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active learning in higher education*, *5*(1), 87-100. doi:10.1177/1469787404040463.

Gómez, G. R., Sáiz, M. S. I., & Jiménez, E. G. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de investigación en educación*, *11*(2), 198-210.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. doi:10.3102/003465430298487.

Hewson, M. G. (2000). A Theory-based Faculty Development Program for Clinician Educators. *Academic Medicine*, *75*(5), 498-501.

Higgins, M., Grant, F., & Thompson, P. (2010). Formative assessment: balancing educational effectiveness and resource efficiency. *Journal for Education in the built Environment*, *5*(2), 4-24. doi:10.11120/jebe.2010.05020004.

Ibarra Sáiz, M. S., & Rodríguez Gómez, G. (2010). Aproximación al discurso dominante sobre la evaluación del aprendizaje en la universidad. *Revista de educación*, 351, 381-407.

Jay, J., & Owen, A. (2016). Providing opportunities for student self-assessment: the impact on the acquisition of psychomotor skills in occupational therapy students. *Assessment & Evaluation in Higher Education*, *41*(8), 1176-1192. doi:10.1080/02602938.2015.1071317.

Jessop, T., & Maleckar, B. (2016). The influence of disciplinary assessment patterns on student learning: A comparative study. *Studies in Higher Education*, *41*(4), 696-711. doi:10.1080/03075079.2014.943170.

Lipnevich, A., & Panadero, E. (2021). A Review of Feedback Models and Theories: Descriptions, Definitions, and Conclusions. *Frontiers in Education, 6*, Article 720195. doi:10.3389/feduc.2021.720195.

Lipnevich, A. A., Panadero, E., Gjicali, K., & Fraile, J. (2021). What's on the syllabus? An analysis of assessment criteria in first year courses across US and Spanish universities. *Educational Assessment, Evaluation and Accountability*, *33*(4), 675-699. doi:10.1007/s11092-021-09357-9.

McMillan, J. H., & Moore, S. (2020). Better being wrong (sometimes): classroom assessment that enhances student learning and motivation. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *93*(2), 85-92. doi:10.1080/00098655.2020.1721414.

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers college record*, *115*(3), 1-47. doi:10.1177/016146811311500307.

Nieminen, J. H. (2020). Disrupting the power relations of grading in higher education through summative self-assessment. *Teaching in Higher Education*, 1-16. doi:10.1080/13562517.2020.1753687.

Norton, L., Aiyegbayo, O., Harrington, K., Elander, J., & Reddy, P. (2010). New lecturers' beliefs about learning, teaching and assessment in higher education: the role of the PGCLTHE programme. *Innovations in Education and Teaching International*, *47*(4), 345-356. doi:10.1080/14703297.2010.518426.

Norton, L., Richardson, T. E., Hartley, J., Newstead, S., & Mayes, J. (2005). Teachers' beliefs and intentions concerning teaching in higher education. *Higher education*, *50*(4), 537-571. doi:10.1007/s10734-004-6363-z.

Palacios, A., and V. M. López-Pastor. 2013. "Haz lo que yo digo pero no lo que yo hago: sistemas de evaluación del alumnado en la formación inicial del profesorado [Do I say but not I do: Students Assessment Systems in Pre-service Teacher Education]." *Revista de Educación 361*. 279–305.

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education*, *44*(3), 379-397. doi:10.1080/02602938.2018.1512553.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74-98. doi:10.1016/j.edurev.2017.08.004.

Polanco-Bueno, R. (2013). Blogs, webinars and significant learning: a case report on a teacher training program for college teachers. *Higher Learning Research Communications*, 3, 56–67. https://doi.org/10.18870/hlrc.v3i1.72.

Postareff, L., Lindblom-Ylänne, S., & Nevgi, A. (2007). The effect of pedagogical training on teaching in higher education. *Teaching and teacher education*, *23*(5), 557-571. doi:10.1016/j.tate.2006.11.013.

Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, *36*(4), 479-492. doi:10.1080/02602930903512883.

Quesada-Pallarés, C., Fernández-de-Álava, M., & Gairín, J. (2017). ¿Cómo aprende el profesorado universitario español? Comprendiendo el uso de estrategias de aprendizaje: How do Spanish university teachers learn? Understanding the use of learning strategies. *Revista de educación*, 376, 135-162.

Rawer, H., Morton, A., McCulloch, R., Heyes, L., & Ryan, J. (1997). Delivering training in teaching skills to hospital doctors. *Medical Teacher*, *19*(3), 209-211. doi:10.3109/01421599709019385.

Raza, S. A., Qazi, W., Khan, K. A., & Salam, J. (2021). Social isolation and acceptance of the learning management system (LMS) in the time of COVID-19 pandemic: an expansion of the UTAUT model. *Journal of Educational Computing Research*, *59*(2), 183-208. doi:10.1177/0735633120960421.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & evaluation in higher education*, *30*(2), 175-194.

Sánchez Moreno, M. R., & Mayor Ruiz, C. M. (2006). Los jóvenes profesores universitarios y su formación pedagógica. Claves y controversias. *Revista de educación, 339, 923-946.*

Steinert, Y., Mann, K., Anderson, B., Barnett, B. M., Centeno, A., Naismith, L., ... & Dolmans, D. (2016). A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. *Medical teacher*, *38*(8), 769-786. doi:10.1080/0142159X.2016.1181851.

Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Gelula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Medical teacher*, *28*(6), 497-526. doi:10.1080/01421590600902976.

Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational research review*, *5*(1), 25-49. doi:10.1016/j.edurev.2009.07.001.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research*, *68*(3), 249-276. doi:10.3102/00346543068003249.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Dordrecht: Springer.

Trowler, P., & Bamber, R. (2005). Compulsory higher education teacher training: Joined-up policies, institutional architectures and enhancement cultures. *International Journal for Academic Development*, *10*(2), 79-93. doi:10.1080/13601440500281708.

Van Zundert, M., Sluijsmans, D., & Van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and instruction*, *20*(4), 270-279. doi:10.1016/j.learninstruc.2009.08.004.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, *10*, 3087.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education*, *45*(4), 477-501. doi:10.1023/A:1023967026413.