ASSESSMENT DESIGN IN HIGHER EDUCATION:

AN ANALYSIS OF DISCIPLINES, TEACHERS' DECISION

MAKING AND TRAINING PROGRAMMES

Universidad Autónoma de Madrid

Author

Javier Fernández Ruiz

Supervisors

Ernesto Panadero

Daniel García Pérez

A mis padres, Guille y Marina…

Agradecimientos/Acknowledgements

Debo decir que me lo he pasado genial durante mis años como doctorando. Puede que el final se me haya hecho un poco cuesta arriba, pero si tuviera que hacer una valoración de los últimos cuatro años, diría que he sido enormemente feliz.

Gran parte de la culpa la tienen las personas de las que me he rodeado, cuyo apoyo incondicional ha sido el mayor regalo que podía esperar.

Empezó con mis padres, Ana e Ismael, en aquel verano de 2017 en el que me mostraron que, aún sin beca, estaban dispuestos a financiarme una tesis doctoral, dos, o las que hicieran falta. Su amor por la educación es contagioso, y creo que al final es el motivo por el que me dedico a esto. Como contrapunto siempre ha estado mi hermano Guille. Dispuesto a debatir constantemente, sobre educación o cualquier otra cosa en la que alguno de los dos creyera tener razón. Intentando ejercer de hermano mayor y enseñarle algo, resulta que al final aprendí muchísimo.

En 2017 empecé mi tesis, supervisado por Ernesto y Dani. No soy capaz de explicar con palabras cuánto he aprendido de vosotros, y cuánto agradezco vuestra mentoría, cariño y tiempo. No he dudado ni por un momento de lo afortunado que soy por poder compartir esto con vosotros. Además, debo dar gracias infinitas a Elena Martín, por acceder a ser mi tutora, y por la inmensa inspiración que ha sido para mí desde que tuve el privilegio de asistir a sus clases en el máster.

No puedo hablar de estos cuatro años sin mencionar a todas las amigas que me han acompañado cuando teníamos algo que celebrar, algo de lo que quejarnos, o incluso ambas cosas. Gracias a Maider y María por todas las noches en el bar de Luis. Gracias Luis por darme una segunda casa, ya que estamos. Y mil gracias a Nano y Vicky por seguir mis progresos y alegrarme con los suyos desde la distancia.

Algo más de dos años pasaron así, hasta que en 2020 tuve la suerte de irme de estancia. I spent about two months in Melbourne, with the amazing researchers of CRADLE. I want to thank all of you for the support and the great memories during my time there. David, Phil, Rola, Margaret, Joanna, Juan and Lasse. All of you made my time there one of the greatest experiences of my life. Thanks also to Gavin, who invited me to Auckland and offered me more support than I could ask for. It was a pity that COVID cancelled our plans, but I hope to visit you in the future.

Si hablamos del tiempo que he pasado fuera de Madrid, necesito mencionar también el año maravilloso que pasé en Bilbao. Allí conocí a mis nuevos compañeros en ERLA, que desde la llegada de Ernesto estaba creciendo a velocidades asombrosas. Resultó que, una vez más, me vi rodeado de gente extraordinaria. Gracias a Ana y Fon, por acogerme en Bilbao. Gracias a Leire, Jose Carlos, Alazne y David por el equipo predoc tan fantástico que hemos formado. Gracias igualmente a Fernando y Eneko por hacerme sentir como en casa, y gracias, una vez más, a Ernesto por invitarme, y por todo en general. Si tuviera que mencionar todo lo que has hecho por mí, este apartado sería más largo que el resto de la tesis.

Por último, pero no menos importante, quiero darte las gracias a ti, Marina. Te escribo directamente porque sé que vas a leer esto, y es que, para mi sorpresa, he descubierto que serías capaz de leerte varias veces mi tesis doctoral sólo para ayudarme a encontrar erratas. Eres una persona excepcional, y me siento increíblemente afortunado de poder compartir mi vida contigo. Gracias por estar a mi lado, por empujarme y por hacerme ser mejor persona.

Gracias por hacerme feliz.

Resumen

Hoy en día está asumido que el diseño de unos métodos de evaluación adecuados es un requisito imprescindible para conseguir una educación de calidad. Sin embargo, los métodos de evaluación que se emplean en las universidades españolas están alejados de las recomendaciones basadas en la evidencia empírica. Muchas de las decisiones relacionadas con el diseño de métodos de evaluación recaen en los docentes universitarios, y comprender el modo en que se toman estas decisiones podría ser de gran beneficio para reducir la distancia entre la investigación educativa y las prácticas reales. Por ello, esta tesis se realiza con el objetivo de comprender cómo los docentes universitarios diseñan sus métodos de evaluación, así como las causas de este diseño. Para ello, se plantean tres estudios.

El primer estudio tiene como objetivo comprender qué efectos tienen las variables externas, en concreto la disciplina de conocimiento, sobre el diseño de evaluación. Para ello, se comparó el diseño e implementación de los métodos de evaluación en tres grados universitarios (ciencias de la actividad física y el deporte, matemáticas, y medicina) en cuatro universidades. Se llevó a cabo una metodología mixta, basada en el análisis de las guías docentes de cada grado (N = 385) así como entrevistas semi estructuradas con los docentes (N = 19). Los resultados muestran la existencia de diferentes tendencias en el diseño de métodos de evaluación en función de la disciplina, divididas principalmente en dos ejes: los propósitos sumativos o formativos de la evaluación, y la evaluación basada en contenidos o la evaluación auténtica.

El segundo estudio tiene como objetivo comprender qué efectos tienen las variables internas, en concreto los procesos de toma de decisiones, sobre el diseño de evaluación. Para este estudio, se realizaron sesiones individuales con parte de los participantes del estudio anterior (N = 16), incluyendo un protocolo en voz alta en el que debían diseñar unos métodos de evaluación para una asignatura ficticia. Las acciones de los docentes durante la tarea fueron transcritas y categorizadas utilizando análisis de contenido y codificaciones abiertas. Se identificaron tres

patrones diferentes: (a) enfocado en la viabilidad de los métodos de evaluación, (b) en el alineamiento con las competencias y resultados de aprendizaje, o (c) en el alineamiento con los métodos instruccionales. La mayoría de los participantes se centraron únicamente en uno de estos elementos. Los docentes también diseñaron métodos diferentes durante la tarea, si se comparan con los que utilizan es sus asignaturas, a pesar de que las competencias y resultados de aprendizaje a evaluar eran los mismos. Los participantes declaran que la falta de recursos es la razón principal de estas diferencias.

El tercer estudio tiene como objetivo analizar el modo en que los docentes universitarios son formados en temáticas de evaluación. Para ello, se recopiló información sobre los cursos de formación docente de todas las universidades públicas españolas (N = 1627), y se analizó el contenido de todos aquellos cuya temática principal fuera la evaluación (N = 82). Se encontraron diferentes enfoques en términos de formato y duración. Mientras algunas universidades ofrecen una gran cantidad de sesiones online, otras prefieren cursos más largos e intensivos. En relación a los contenidos de los cursos, se identificaron 25 temas, agrupados en 6 áreas temáticas. El formato y los contenidos de los cursos fueron comparados con la literatura acerca de formación docente en evaluación. Por último, se propusieron tres criterios de calidad basados en los modelos teóricos de formación docente en evaluación, encontrando que sólo 3 de las 25 universidades analizadas ofrecerían una formación en evaluación de alta calidad a sus profesores.

Los tres estudios propuestos en esta tesis ofrecen una perspectiva más amplia acerca del modo en que los docentes diseñan sus métodos de evaluación, y de las herramientas que necesitan para hacerlo. Con el propósito de integrar los resultados de los tres estudios, en la discusión se proponen dos modelos basados en la evidencia acerca del diseño de evaluación. El primero, Variables influencing assessment design (VIAD) se realiza una clasificación y agrupación de las variables que condicionan el diseño de métodos de evaluación en los profesores universitarios. En el segundo, Teachers' Assessment Design and Implementation Process (TADAIP) se ofrece un modelo secuencial que muestra las fases por las que un docente pasa al diseñar sus métodos

de evaluación. Además, se discuten las implicaciones para la práctica profesional y para investigaciones futuras extraídas de esta tesis.

Abstract

Today it is assumed that the design of adequate assessment methods is an essential requirement to achieve a quality education. However, the assessment methods used in Spanish universities are far from the recommendations based on empirical evidence. Many of the decisions related to the design of these methods are in the hands of university teachers and understanding how their decisions are made could be of great benefit in reducing the gap between educational research and practice. Therefore, this thesis is carried out with the aim of understanding how university teachers design their evaluation methods, as well as the causes of this design. For this, three studies are proposed.

The first study aims to understand the effects of external variables, specifically the discipline, on assessment design. To do so, assessment design and implementation in three programmes (sport science, mathematics, and medicine) each representing a discipline from 4 Spanish universities was compared. Using a mixed-methods approach, data from syllabi (N = 385) and semi-structured interviews with teachers (N =19) were analysed. The results showed different approaches to assessment design and implementation in each programme in two main axes: summative or formative purposes of assessment, and content-based or authentic assessment.

The second study aims to understand the effects of internal variables, specifically decision-making processes, on assessment design. Sixteen teachers from four universities participated carrying out a think-aloud simulation task: designing the assessment methods for a set of learning outcomes. Teachers' testimony during the task was transcribed and categorized using content analysis and an open-coding procedure. Three different patterns were identified: (a) focus on the feasibility of the assessment tasks, (b) on the alignment with the learning outcomes, or (c) alignment with teaching methods. Most of the participants focused only in one of the three elements. Teachers also designed different assessment methods in the simulation task in comparison with the ones they use in their

subjects, despite the guidelines of assessing the same learning outcomes. A lack of resources is claimed as the reason of these differences.

The third study aims to understand how university teachers are trained to design and implement assessment. To do so, every teacher training course from Spanish public universities (N = 1627) was screened. Data about all available courses related to assessment (N = 82) was collected and analysed. Different approaches in terms of format and duration were found depending on the university. While some universities use a massive webinar approach to teacher training, others prefer longer and more intensive courses. Regarding courses' contents, 25 themes were found, grouped into 6 main thematic areas. Courses format and contents were compared with literature foundations on assessment training. Lastly, three quality criteria are proposed based on the theoretical models about assessment literacy, finding that only 3 of the 25 universities analysed would be offering a high-quality assessment training to their teachers.

The three studies proposed in this thesis offer a broader perspective on how teachers design their assessment methods, and the tools they need to do so. In order to integrate the results of the three studies, the discussion proposes two models based on evidence about the assessment design. The first, Variables Influencing Assessment Design (VIAD), is a classification and grouping of the variables that condition the design of assessment methods for university teachers. In the second, Teachers' Assessment Design and Implementation Process (TADIP), a sequential model is offered that shows the phases a teacher goes through when designing their assessment methods. In addition, the implications for professional practice and for future research drawn from this thesis are discussed.

Index

Table index

Figure index

CHAPTER 1

My academic path: What I have learned and why I am doing this thesis

This doctoral thesis might be better understood if I start by telling my history at ERLA group, and the moments that have brought me here. I started working with ERLA in 2016, and several of the things I have worked on, or what I have learned are directly reflected in this doctoral thesis. The next figure covers the collaborations with more influence in the development of this thesis.
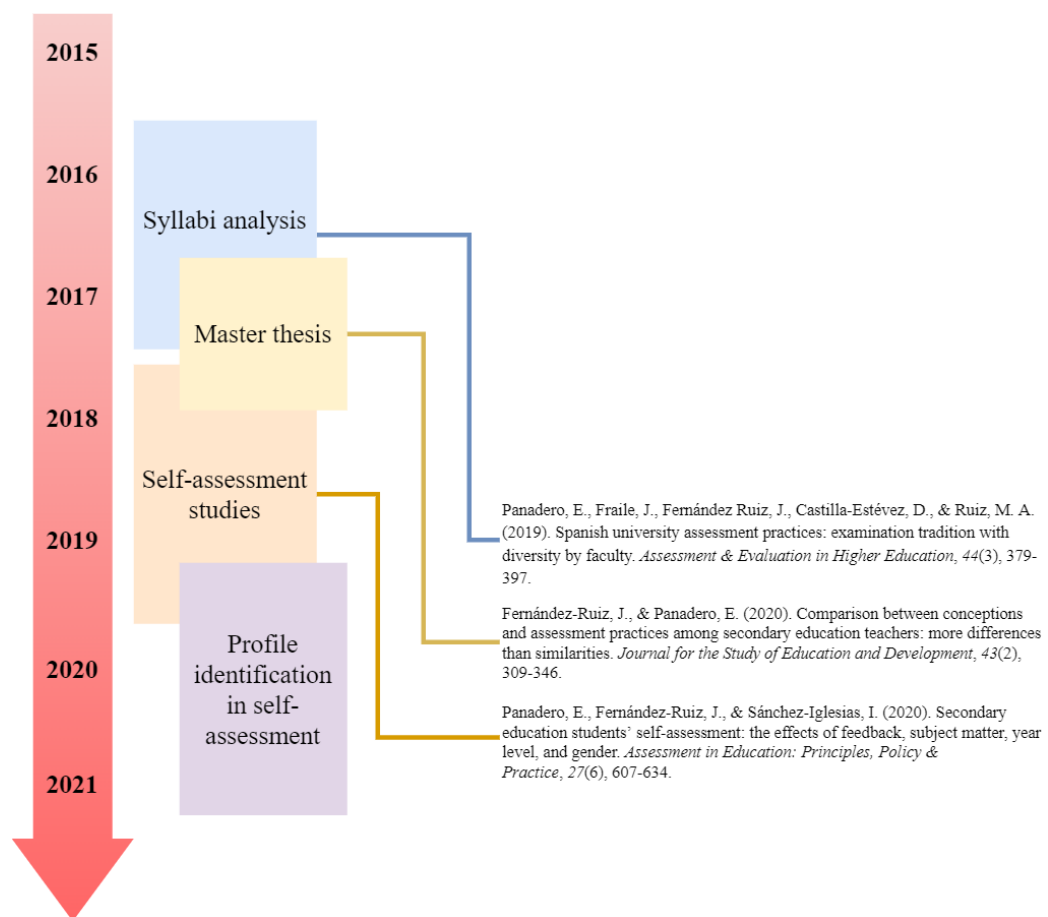
Figure 1.1. Key influences during my activity on the team.

To begin with, the first of my tasks as a young researcher consisted of counting the syllabi from all Spanish public universities. We discovered that there were around 70,000 teaching guides in the 50 Spanish public universities and that, if we needed a reasonably representative sample, we had to reach at least 2% of the total guides. So, my next job as a researcher in the team was to collaborate in the analysis of a large part of the 1,500 guides that we needed.

The results of this study really impressed me. They showed that despite the improvements during the last decades, the assessment methods carried out in higher education were still far from being as the literature recommended it to be. I was influenced by these results, up to the point that my doctoral thesis is a continuation of this line of research. If this study investigated what assessment methods are carried out in Spanish higher education, my thesis investigate why teachers are using these methods.

This research question could have been approached from different levels, but in my case, there was another research experience that had made me choose a particular perspective. This research experience was my masters' thesis where I investigated the assessment conceptions of secondary education teachers and how these are reflected in the assessment practices they use in the classroom. Working on this study I understood that teachers are a central agent in the assessment processes, and that their perspective is enormously complex and deserves to be investigated in detail.

The two mentioned studies are the ones that have influenced the topic of my doctoral thesis. However, the influences that my previous work in ERLA reach many other levels. To start with, the data collection methods. During the syllabi study I also learned that educational research does not have to be exclusively linked to human samples, but that there is a huge number of public documents related to the educational legislation and practices that contain important data that has not yet been analyzed. This fact has led me to include document analysis as an important part of this doctoral thesis.

There has also been another study that has influenced the data collection methods used in this thesis. This time the study was launched in 2018, and it was aimed at understanding the way students performed self-assessment. It was a large-scale investigation in which we interviewed almost 200 high school and university students, and, once again, in my role as a young researcher, I actively collaborated in these

interviews. There I realized several things that later were reflected in my thesis. The first is the importance of a semi-structured interview as a method of acquire a deep understanding of a participant point of view. I was intrigued by the participants' testimony, and decided that, aiming to understand how university teachers design assessment, I would need to interview them.

The self-assessment study included another data collection method: think-aloud protocols. This way eliciting cognitive processes was totally new to me. I had not read much about how it could be used, what was its usefulness and what differentiated it from self-report methods. Both data collection methods included in this study helped me to understand, while conducting the interviews, how important data triangulation was. Therefore, I decided that I could not rely solely on self-report methods. That is why this thesis has also included think-aloud protocols.

When we finished the self-assessment data collection process, we analyzed the data and wrote several papers. Some of them are already published. However, we also concluded that our data was too rich to be limited to quantitative analysis. It is at this moment that our team started the last endeavour that has directly influenced the development of my doctoral thesis, in this case related to data analysis methods.It was decided that it would have great scientific interest to propose a study about the existence of differentiated profiles in the way in which students perform self-assessment. Once again, I ended up collaborating in this new approach to data analysis. I found it so interesting, so rich and innovative that I decided once again that I wanted something like that to be included in my doctoral thesis.

As we can see, this thesis can not be understood without my experiences from ERLA. Previous studies in which I have collaborated have directly influenced this thesis

both at the level of topic and agents to study, data analysis, and data collection methods.

Next, we will proceed to present this thesis.

CHAPTER 2

Educational assessment: A powerful instructional and learning strategy

"The single, strongest influence on learning is surely the assessment procedures (...) even the form of an examination question or essay topics set can affect how students study (...) It is also important to remember that entrenched attitudes which support traditional methods of teaching and assessment are hard to change."

(Entwistle, 1996, pp. 111–112).

Assessment research: then and now

Educational assessment is one of the most studied topics in educational research (Bennet, 2011), currently involving many lines of study, such as the purposes (Black & Wiliam, 1998), components (Hattie & Timperley, 2007), agents (Yan, 2020) or outcomes of assessment (Yin et al., 2008). As with any other research field, assessment has significantly developed over decades of study. Educational assessment, originally, was considered as a way of quantifying student learning and the effectiveness of systems with diagnostic purposes (Brown, 2018). However, in the second half of the 20th century, this vision of solely diagnostic assessment began to progressively change.

Scriven (1967) coined the term "formative assessment" and applied it to educational programmes, distinguishing assessment aimed at evaluating the programme quality from that aimed at improving programmes. Bloom (1969) used this same distinction, but applied it to students, which is remains the most common use of the term. The seminal paper by Black and Wiliam (1998) helped to expand the concept of formative assessment into a more fully elaborated concept. In their vision of formative assessment, they explored empirical and theoretical aspects of the assessment process that were not yet sufficiently salient in the general education literature, including the complexity and potential of feedback or students' participation in assessment via self- and peer-assessment. Afterwards, strong new lines of research on formative assessment erupted to

explore aspects related to student learning, such as the effects of assessment methodologies on self-regulation (Clark, 2012; Nicol & Macfarlane-Dick, 2006), motivation (Cauley & McMillan, 2010) or emotional reactions (Värlander, 2008), all of which began to be empirically analysed.

The revolution started by, among others, the work of Black and Wiliam during the 1990s and the beginning of the 21st century brought with it a new way of understanding assessment as a method for improving student learning. Nevertheless, educational systems continue to use assessment as a method of measuring and accrediting student achievement via grading, with grades being used to take decisions regarding promotion at all educational levels. It is therefore important to explore in more detail the relationship between the summative and formative purposes of assessment.

Formative and summative assessment

There are many decisions that must be made regarding assessment, the first of which is defining its purpose. Depending on the aim pursued, the assessment methods used will follow very different approaches. The literature has extensively studied the purposes of educational assessment, highlighting the distinction between formative and summative assessment.

Formative assessment is considered a critical component of effective instructional practice (Pellegrino et al., 2001; Ruiz-Primo, 2011). Black and Wiliam (2009) define formative assessment as:

> A class practice is formative when the evidence about the student's performance is made explicit, interpreted and used by the teacher, student or their peers to make decisions about the next steps in instruction that will be better, or better supported, than those they would have taken in the absence of the evidence that was made explicit. (Black & Wiliam, 2009, p. 9)

Summative assessment, on the other hand, is defined as "cumulative assessments (…) that intend to capture what a student has learned, or the quality of the learning, and judge performance against some standards" (National Research Council, 2001, p. 25). However, the purposes of assessment are more complex than a mere dichotomy between formative or summative assessment. Brown (2018) raises three questions to help define the purpose of assessment:

1. Is assessment useful?

2. If it is useful, when does assessment take place in the educational process?

3. Independent of timing, is assessment aimed at informing improvements in the processes or at evaluating the quality of the processes?

The first question explores the possibility that assessment is useless, meaningless, and even counterproductive. This view of assessment tends to be little discussed, although it is a legitimate concern, as we know well from the review by Kruger and DeNisi (1996), which found that in one third of the studies they examined, feedback had a negative impact on subsequent performance. There are teachers who believe that the efforts carried out in educational assessment are not worthwhile, considering the usefulness of the information obtained from it (Brown, 2004).

The second question explores the importance of time for assessment purpose. If it is intended for carrying out a formative assessment, which encourages and promotes student learning, it cannot be done at the end of the process (Black & Wiliam, 1998). It is necessary to offer students the opportunity to extract information about their performance and to use this information to achieve better results. In the case of summative assessment, if what is intended is to measure learning outcomes, for students or for the school's accountability purposes, it is necessary that these measurements are carried out

at the end of the process, to obtain a definitive measurement of the development of students' knowledge and skills.

Although there is a general consensus about the usefulness of assessment, the opinions about whether it should be used for improvement or accountability – and the ways to achieve each purpose – are still mixed. Debate about the purposes of assessment has been present in the literature for several decades (Lau, 2016), since the publication of seminal works such as Black and Wiliam's (1998), which addresses the importance of formative assessment, or Biggs (1998), which critiques Black and Wiliam's review for excluding the effects of summative assessment. However, in recent years, a trend has emerged in favour of formative assessment as the way to go for educators and policymakers. This trend is reflected in the literature to the point of formative assessment being portrayed as "good" assessment, while summative assessment would be "bad" (Taras, 2005).

As a counterpoint, other authors such as Scriven (1991) or Brown (2018) subscribe to the idea that formative and summative assessment have the same characteristics and must be of the same quality. The difference between them would just be the time when they are carried out. In this approach, formative and summative assessment are not treated as different practices, but as the same practice carried out at different times. Lau (2016) argues that formative and summative assessment need to coexist and be connected with each other, as well as with the overall learning and teaching environment. This is especially important in the educational level explored in this dissertation, as all students are graded at the university. In short, it is necessary to go beyond the simple idea "formative good, summative bad".

The debate between formative and summative assessment is not as active in educational practice, or at least is much subtler. The vast majority of teachers, especially

in secondary and higher education, tend to use a combination of formative and summative assessment because, in real settings, it is more difficult to differentiate them. Additionally, on many occasions, the teachers are not aware of the theoretical development regarding assessment purposes. The case of higher education is particularly interesting. Although traditionally an eminently summative assessment was offered, in the form of an exam or test at the end of the subject (Zabalza, 2003), after the implementation of the Bologna plan in European universities, continuous assessment has been promoted throughout the subject, trying to provide a formative impulse to assessment practices (Panadero et al., 2019). This impulse requires a lot of coordination and involves several agents. Teachers are one such agent and their decisions are especially relevant to create change in the instructional and evaluative practices offered.

It is in the hands of teachers to implement assessment methods which cover the requirements depending on the purpose of the assessment, and to do so, there is a vast range of decisions they must make. Many of those decisions are related to the construction and balance of the different assessment components.

Components of Assessment

In any case, whether it is an assessment with formative or summative purposes (or mixed, as usually happens), it is evident that assessment is a complex process with several issues to consider. These issues will be mentioned several times throughout this thesis. Below we offer a definition of the six components of assessment that are most relevant for our subsequent studies.

Assessment evidences. An assessment evidence is the tangible product chosen by the teacher for students to demonstrate their learning. Different authors, such as Carless (2015), García-Pérez, Fraile & Panadero (2021) and Sambell et al. (2013) pay particular attention to assessment evidences as powerful instruments to help students acquire deep

approaches to learning. In higher education, the most widely used assessment task tends to be final exams (Lipnevich et al. 2020). However, in recent years there has been a tendency to complement this evidence (eminently summative) with others such as assignments, classroom practices, or portfolios (Panadero et al. 2019). Students, however, have their own perspectives about assessment evidences. According to Ibarra Sáiz et al. (2020), students prefer tasks to be challenging, eminently practical, and connected with professional reality. As authors propose (Rodríguez-Gómez & Ibarra-Sáiz, 2015), to achieve assessment tasks in line with students' needs, requires a change in the mentality in both educators and students.

Assessment criteria. Specific assessment criteria tend to be defined explicitly or implicitly for each task (Sadler, 2005). These criteria allow the teacher to make decisions about the students' level of performance, which may ultimately have accrediting effects on the student's academic path. Although the assessment criteria are usually poorly shared with the student (O'Donovan et al., 2004), there are methodologies that help make them clearer and understandable for all parties, such as the use of assessment rubrics (Andrade et al., 2008).

Grading. Grading is an integral part of the instructional process. Many formal educational systems make it compulsory for the teacher to make accrediting decisions about the performance of their students, and higher education does not escape this trend. Grades provide useful information to all of the agents involved in the assessment process, certifies the attainment of learning goals and provides a basis for improvement efforts. Good grading, however, requires well-designed assessment tasks. As Guskey and Bailey (2001, p. 2) argue, "even the most sophisticated, technologically advanced reporting system cannot compensate for poor, inadequate, or unreliable evidence on student learning". Validity of grades has been extensively studied in relation with several

variables. Researchers have described grades' connection with multidimensional measures such as motivation, engagement, and persistence (Willingham et al, 2012). Another group of recent studies measure the relationship between grades and other non cognitive variables, such as academic success (Cliffordson, 2008; Bowers, 2014) or academic dropouts (Gleason & Dynarski, 2002).

Feedback. The information that students receive about their own performance in assessment tasks is considered feedback (Hattie & Timperley, 2007). Feedback is one of the most decisive factors in student learning, and its characteristics and effects are widely studied in the scientific literature (Lipnevich & Smith, 2009, 2018; Winstone & Boud, 2018). The challenges of providing effective feedback have been well discussed in several recent articles (Carless et al., 2011; Henderson et al., 2019). Among others, students often view comments by teachers on their work as difficult to understand (e.g. Weaver, 2006), lacking specific advice on how to improve their own personal work (e.g. Higgins et al., 2001) or difficult to act upon (e.g. Gibbs, 2006; Poulos & Mahony, 2008).

Agents. Based on formative assessment principles, there has been a trend in educational research and practice that moves the teacher away from holding the monopoly of assessor roles in favour of student participation (Black & Wiliam, 1998; Falchikov & Boud, 1989). There are several ways to ensure the student actively participates in assessment methods, either by discussing assessment criteria with the teacher, jointly designing rubrics (Fraile et al., 2017), or through self-assessment or peer assessment. These last two practices place students as evaluating agents of their own work and that of their peers, and their pedagogical potential and their effects on a large number of variables are being widely studied in recent scientific literature (Andrade & Valtcheva, 2009; Brookhart & Chen, 2015; Li et al., 2020; Yan, 2020).

Educational assessment occurs at all educational levels, as well as in a wide variety of other instructional contexts such as university entry level exams, MOOCS or professional certifications. However, the focus of this dissertation is assessment in higher education. This distinction is relevant as assessment in higher education has several characteristics which makes it a unique context (e.g. students more capable of self-regulating, different training regarding teachers' assessment literacy). In the next section, we dig into research concerning educational assessment in the university context.

Assessment in Higher Education

One major agreement regarding assessment in higher education is that, although there have been recent advances, the assessment practices carried out still have room for improvement, such as enhancing the formative purposes (Lipnevich et al., 2021; Panadero et al., 2019). The literature traditionally shows a predominance of the final assessment, or assessment of the product (Ibarra-Saiz & Rodríguez-Gómez, 2010; Zabalza, 2003), over the continuous assessment of the students, or assessment of the process. This is counterproductive for students' learning, because it limits their opportunities to receive feedback and improve their performance before delivering the final product (Hernández, 2012). However, it is worth highlighting the efforts to promote continuous assessment in European higher education by supranational educational plans such as the Bologna plan (Wätcher, 2004). The effects of these efforts can be seen in the timid increase during the last ten years of continuous assessment in universities in some European countries (Llamas et al., 2018) and the increase in the number of assessment practices used (Panadero et al., 2019).

Another concern is the scarce participation that students have in university assessment practices. Students of this level stand out for their capacity for analysis and self-regulation and would clearly be possible to involve them more actively in the

assessment processes, either through techniques such as self- or peer assessment, or in the construction of assessment tools such as rubrics (Fraile et al., 2017). Assessment methodologies promoting students' participation, such as self- or peer assessment, have been discovered as a useful method to increase self-regulation (Panadero & Romero, 2014), motivation (McMillan & Hearn, 2008) and achievement (Brown & Harris, 2013; Yan, 2020). However, their use is very limited in universities (Fraile et al., 2018; Panadero & Brown, 2017).

Finally, research has also shown important differences between disciplines in relation to the assessment methods chosen (Lipnevich et al., 2020; Panadero et al., 2019). In this sense, there are plenty of studies exploring each discipline on their own (e.g. physical activity; López-Pastor et al., 2013; Schuwirth & Van der Vleuten, 2019; Trenholm et al., 2015), but there is a lack of comparative and general studies that cover several academic areas at the same time. An exception is the work by Jessop and Tomas (Jessop & Tomas, 2017; Tomas & Jessop, 2019), Lipnevich et al. (2020, 2021) and Panadero et al. (2019). The main finding from these studies is that there are important variations among disciplines and, therefore, we need to study those in more detail. Following this research logic, this dissertation aims to compare different undergraduate programmes.

As part of assessment in higher education, this research is also intensely focused on one particular assessment agent: the teacher. The teacher is, in most cases, the ultimate person responsible for choosing and developing a specific set of assessment methods. Accordingly, every intervention aimed to raise the quality of assessment in higher education must acknowledge the role of teachers. First, we expand on how assessment design works in real practice for university teachers.

Assessment design

Institutional Requirements for Assessment Design

Educational assessment is a complex phenomenon. If correctly designed and implemented, it can quantify students' learning, leading to more effective and adjusted diagnostic or accrediting decisions, while also enhancing learning at the same time. Although there is an enormous amount of literature on the purposes, components and outcomes of assessment, the research shows that the methodologies implemented in higher education do not always reflect the theoretical knowledge available on the subject (Ibarra-Sáiz & Rodríguez-Gómez, 2010). University teachers are using traditional assessment methodologies and have problems clarifying the purpose of their assessment methods (Postareff et al., 2012). To address this problem, it is necessary to understand how the process of design and implementation of assessment methods works in higher education.

Next we will present the most typical route to the design of assessment in the Spanish university system, the context in which this dissertation takes place. Before starting the academic year, sometimes in the middle of the previous one, teachers are required to write the syllabus for the course. A syllabus is an official document in which the teacher of a course provides pedagogical information, expectations and responsibilities to students (Stanny et al., 2015). This document is the academic guide for the subject during the following year, and includes information about the contents, the learning outcomes to be achieved by the student and the teaching and assessment methods.

This process of designing assessment can be problematic for teachers due to several factors. First, its early design means that on many occasions they still do not know the number of students they will have in their course, which is common in elective courses

with highly variable ratios from one year to another. In addition, it may be the case that the teacher writes the syllabus for his course without being sure if he or she is going to teach it next year, or that he reaches a new course with a syllabus written by the previous teacher. All of these factors can influence the syllabus ending up, in many cases, as a vague document that only describes the general guidelines regarding assessment methods, although these guidelines are unchangeable throughout the course, and refer in most cases to the tasks used and how they are graded.

For the remaining aspects related to the assessment, such as the feedback received by the students and their participation in the assessment, the teachers are free to make decisions during the academic year, depending on the context and the students' response. However, sometimes these decisions must be agreed upon by the teaching unit or by the department, which is not always an easy task. The process of design and implementation of assessment methods in higher education thus appears to be an extremely complex and context-dependent process. As such, in recent years, a new interest has emerged in the scientific literature to learn about the perspective of teachers throughout the process of designing assessment methodologies.

Summary of Research on Assessment Design

Assessment design is a relatively recent topic within assessment literature. During the last few decades, new articles have been published regarding the way in which teachers design their assessment methods. These articles sometimes referred to specific aspects of teachers, such as professionalism (Norton et al., 2019), decision-making (Bearman et al., 2016) or assessment literacy (Smith et al., 2013). Likewise, articles have been published on the design of assessment methods in specific disciplines such as health or the hard sciences.

However, a general theoretical framework was still lacking. An interesting proposal was published by Bearman et al. (2016, 2017). In their 2016 study and related works, these authors developed the so-called *Assessment design decisions framework*, which was intended to guide teachers in making better decisions about the design of their assessment methods. The framework includes aspects related to the assessment, such as its purpose, the contexts in which it will be carried out, the alignment with the competences of the subject, the degree or future profession of the students, the type of evidence to use, the feedback offered and the interactions with other educational agents. Finally, they made five recommendations regarding the design of assessment methods:

1.   Benefit the learner but support the educator. As the basis for the assessment practices may be learners, educators' circumstances must be considered.

2.   Design is individual but also distributed. Assessment is mediated by individuals but involves collaboration and negotiation with a range of colleagues.

3.   Holistic design processes blend with strategic decisions. Even if the assessment design processes look holistic and intuitive for some teachers, educators need to make strategic choices to correctly develop and implement their assessment methods.

4.   Think conceptually, relationally but also pragmatically. Improving assessment practices requires a balance between conceptual, interpersonal and pragmatic aspects.

5.   Think locally but also beyond the square. Assessment is highly contextualised and influenced by local cultures. However, to innovate it is necessary to transcend these influences and look for alternative perspectives.

Starting from the literature such as the assessment design decisions framework, and aiming to putting it in practice, it is necessary to understand how teachers make decisions, which contextual and personal factors might be influencing these decisions and how can we help them to make better-informed decisions.

Design Decisions

There is a large body of literature on the way teachers make decisions about their work, whether in terms of instructional design or assessment. Since the 1980s, the teaching profession has been understood as one of a markedly reflective nature (Schön, 1984) that requires a constant review of the practices carried out. Authors such as Anderson (2003) explore the ways in which teachers make decisions, providing several insights into how this process may develop.

For example, teachers can decide to continue with their previous practices. It is important to highlight that a decision-making process does not always entail a redesign of the teacher's practices, which can lead to the conclusion that the best way to act is the one currently implemented. However, the teacher can make changes to their methodologies during this decision-making process. Anderson (2003) and Bearman et al. (2017) agree on the possibilities for teachers to make decisions influenced by the requirements of the context in which they carry out their practices. Additionally, teachers can also make decisions based on the information they have collected from their students during the time they have been in practice. This information is known as the "backwash effect" (Watkins et al., 2005) and it has been frequently cited as a relevant factor according to which teachers make decisions.

The decisions made by teachers can be classified into three different categories (Bearman et al., 2016); they can incorporate modifications in essential, selective, or meta-design activities. The first category, referring to essential activities, includes those that design or modify specific assessment evidence. Teachers can select the type of task, the feedback processes associated with it and its alignment with teaching methods or learning outcomes. Second, selective activities are those that include changes or modifications in aspects of the qualification of the previously designed evidence. These activities may be

aimed at making grades fairer for the student, reducing contract cheating or plagiarism or ensuring fairness in the assessment methods used. Third, meta-design activities are those related to decisions about the assessment design itself. These activities may include a periodic review of the assessment methods implemented, the participation of students in the implementation of these methods or collaboration with other teachers.

The results of Bearman et al. (2017) offer very valuable complementary information to understand the results of Panadero et al. (2019) and Lipnevich et al. (2020, 2021) and, along with similar works (Macdonald & Joughin, 2009; Price et al., 2011), allowed the identification of some factors that teachers should consider when designing their assessment methods.

Contextual Factors with a Focus in the Role of Disciplines

Whether there are differences between academic disciplines in the way they assess their students has been recently studied. Studies such as Panadero et al. (2019) or the ones by Lipnevich et al. (2020, 2021) show certain traditions in each of the disciplines associated with different approaches to assessment. Given that these studies were carried out with a sample made up of syllabuses, their results offer us a photograph of the status of the assessment at the beginning of the academic year, but they do not offer information about the design process that led to these methods, nor about the characteristics of their subsequent implementation. To achieve an in-depth vision of how different disciplines approach assessment design and implementation is the main aim of the first study of this thesis.

As a starting point, Panadero's work and similar studies (Lipnevich et al., 2020) offer us an interesting vision of the differences between academic disciplines in terms of assessment methods. It is known that in degrees in the social sciences, especially those related to education, the assessment of learning tends to be more formative in nature than

in other disciplines (Hay, 2006; López-Pastor et al., 2013). This may be the result of the pedagogical knowledge of its teachers, who are more familiar with innovative assessment methodologies.

In the case of the health sciences, it is worth noting the great change in focus that has been experienced in recent years (Govaerts, 2008). The focus on specific content and theory assessment that was characteristic of this discipline is being transformed into assessment methodologies more focused on simulating the professional practice of its students (Schuwirtz & Van der Vleuten, 2019). In the case of STEM-related disciplines, the spectrum is broader, from those focused on theoretical tests above all else, such as mathematics, to those with a more practical approach, such as some engineering fields. However, studies such as Ianonne and Simpson (2011) or Lipnevich et al. (2020) show much greater use of examinations in the hard sciences.

Despite being aware of the differences in the way teachers in each discipline evaluate students, there remains a gap in research regarding the causes of these differences. Studies that explore the design process for assessment methods in different academic disciplines are still scarce, which adds to the innovative value of this doctoral thesis. Another contextual factor that is mentioned frequently in the study by Bearman et al. (2017) is the availability of material or temporal resources. The student ratio and the workloads of the teacher or the students are fundamental factors when deciding on specific assessment methods. Other environmental factors such as interactions with peers or with students, as well as the characteristics of certain student cohorts also significantly affect the way in which teachers decide to evaluate their students.

Individual Factors

Apart from external factors such as discipline, assessment design is also influenced by personal or individual factors. The literature has explored several of these

factors in recent years, helping to clarify how assessment design is also an individual process. Our second study aims to contribute to this line of research by exploring the patterns among teachers regarding the design of their assessment methods. As a starting point, the literature describes several factors that influence teachers in assessment design. The first factor covers conceptions or beliefs about assessment. Brown (2003, 2004, 2006) has carried out extensive work on teachers' conceptions about assessment. It is now known that those conceptions might include the view of assessment as a method for improving learning (in line with the principles of formative assessment), as a method for measuring learning or as a tool for teacher or school accountability (Remesal & Brown, 2015). Teachers' conceptions about assessment are also closely related with the assessment practices they design and carry out. For example, Fernández Ruiz and Panadero (2020) have shown how teachers with formative conceptions of assessment are more willing to use more continuous assessment methods, provide more feedback or implement peer assessment.

The next factor is past experiences and backwash effects. As an example of this line of research, Panadero et al. (2014) and Panadero and Brown (2017) have explored how previous positive experiences with self- and peer assessment and awareness of their advantages make teachers more willing to implement them in their classes. Additionally, teachers analyse students' responses and use this as a measurement of the success of the methods implemented. This is a phenomenon known as the backwash effect (Watkins et al., 2005) and has been admitted as another influence to assessment practices.

Teaching experience is another relevant and personal factor. New teachers tend to have different approaches to assessment than veteran teachers (Quesada-Serra et al., 2016). The differences are especially relevant in terms of formative or summative approaches to assessment, and attitudes towards innovative assessment practices. Novice

teachers tend to consider it more important to promote formative assessment, perform an initial assessment and make students participate in self- and peer assessment compared to their more experienced peers.

Lastly, assessment literacy is also acknowledged as a relevant factor related to assessment design. Looney et al. (2017, p. 2) define assessment literacy as "teacher capabilities to plan and implement quality assessment tasks, to interpret evidence and outcomes appropriate to the assessment purpose and type, and to engage students themselves as active participants in assessment of their own learning". How to promote assessment literacy among teachers has been a scientific concern over the past decades and will be discussed next.

Assessment Literacy

There is abundant literature conceptualising what makes a teacher assessment literate. Aside from the previously mentioned definition, other authors have argued that assessment-literate teachers understand how to construct, administer and score reliable assessments and communicate valid interpretations about student learning (Popham, 2011). In a deeper approach, several authors (DeLuca, 2012; Xu & Brown, 2016; Pastore & Andrade, 2019) have proposed models of assessment literacy. DeLuca's model is based on the ICE (ideas, connections, and extensions) model by Young et al. (2000). It also has a major focus on assessment fairness as an overarching construct in the model.
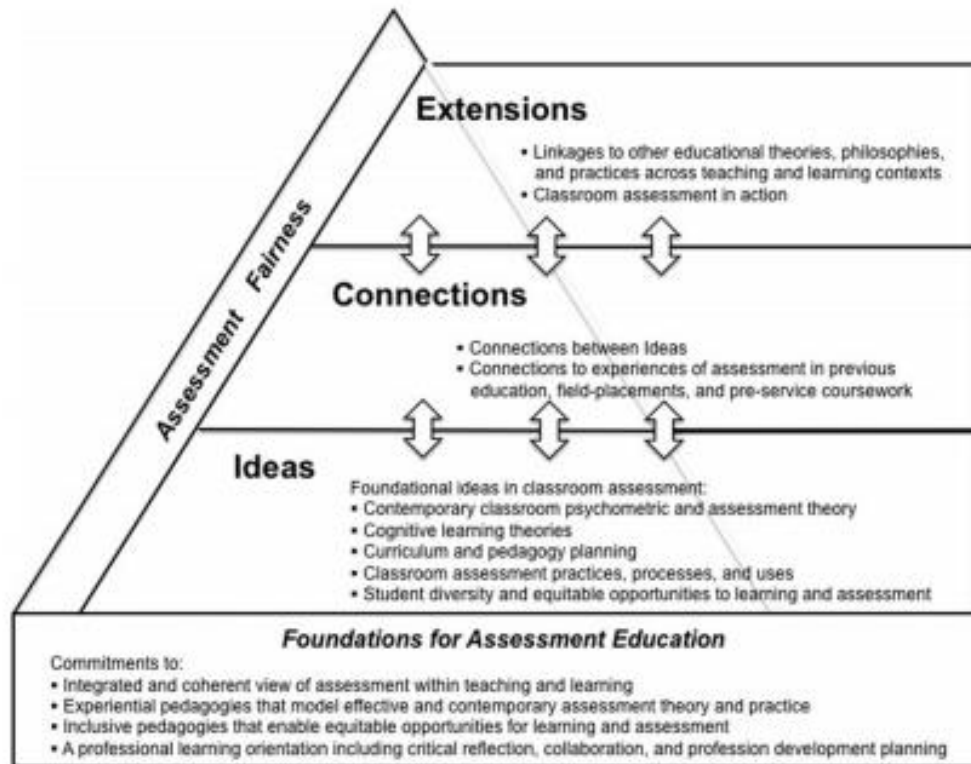
Figure 2.1. Assessment education framework (Deluca, 2012, p. 586)

As shown in figure 2.1, this model is composed of the three classical areas in an ICE model: Ideas, Connections and Extensions. At the first level, teachers would be exposed to the key ideas about classroom assessment based on current research and government standards. Most assessment literacy models include a theoretical knowledge base recommended for teacher training programmes. In this case, DeLuca (2012) recommends classroom psychometric and assessment theory, cognitive learning theory, curriculum and pedagogy planning, educational standards and policies and classroom assessment practices. At the second level, teachers would link these ideas with each other and with their own experiences related to assessment. Teachers can link their knowledge with their previous assessment experiences as students or pre-service teachers. Achieving this type of connection reinforces the ideas about assessment previously mentioned, while also building commitment to using assessment in more constructive and positive ways (DeLuca, 2012). In the final level, the teachers would expand on the assessment concepts

created in the two previous levels, building bridges in the direction of general paradigms of learning. At this point, assessment knowledge is totally integrated into teacher's pedagogical philosophies teaching practices.

DeLuca's ICE model served as the basis for other influential assessment literacy models. For example, Xu and Brown (2016) presented their assessment literacy model, conceptualised as Teacher Assessment Literacy in Practice (TALiP). This model is especially focused on understanding assessment literacy as a phenomenon with practical consequences, instead of a merely cognitive process. As can be seen in figure 2, it as begins with a knowledge base involving several areas (Brookhart, 2011), which would be filtered and interpreted by each teacher.

A conceptual framework of teacher assessment literacy in practice
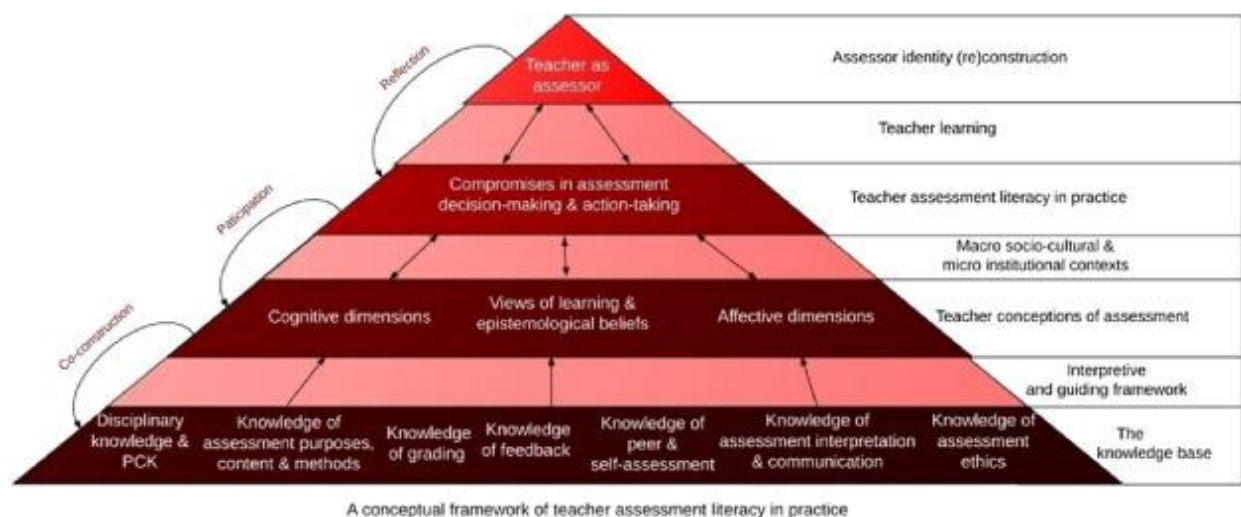
Figure 2.2. Teacher assessment literacy in practice (Xu & Brown, 2016, p. 155)

However, as these authors argue "teachers cannot do whatever they please in actual practice since they are employed within an immediate workplace community and larger social, political, and cultural contexts" (Xu & Brown, 2016, p. 157). Accordingly, the next level of the model would be oriented towards teachers' decision-making. This is a process by which teachers would balance the demands and constraints of those external factors with their own beliefs and values (McMillan, 2003). Teachers need to make compromises, as Carless (2012) has argued, to engage in good assessment practices, as

their conceptions and expectations about assessment are probably not fully suitable to the context. Therefore, Xu and Brown (2016) argue that teacher assessment literacy is better understood as teacher assessment literacy in practice, which consists of various compromises that teachers make to reconcile tensions.

Lastly, Pastore and Andrade (2019) presented a three-dimensional model of assessment literacy, aiming to identify the teachers' assessment needs to best respond to educational policy requirements while also supporting student leaning. These authors consider that an assessment-literate teacher must understand and differentiate the aims of assessment, and also "articulate a cyclical process of collection, interpretation, use of evidence and communication of feedback" (Pastore & Andrade, 2019, p. 135).

Three-dimensional model of assessment literacy
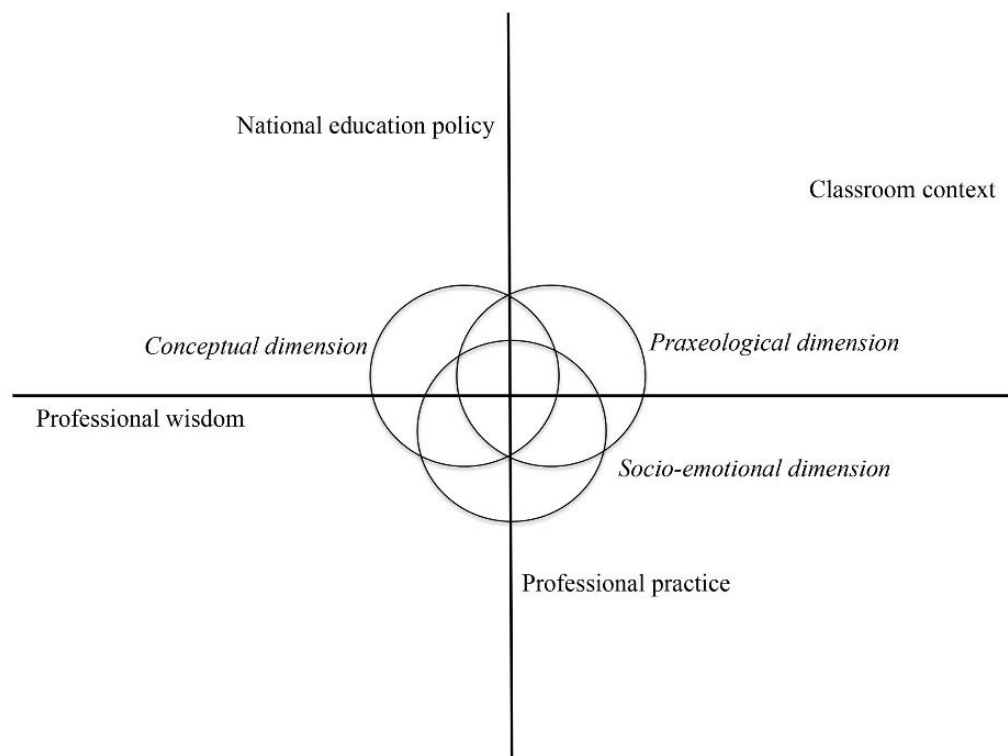
Figure 2.3. Three-dimensional model on assessment literacy (Pastore & Andrade, 2019, p. 135).

Pastore and Andrade's model present a three-dimmensional approach to assessment literacy. First, the model would cover a conceptual dimension, based on what

teachers need to know about assessment, while also acknowledging the filter of each teacher's individual conceptions. Second, the model presents a praxeological dimension exploring assessment as a practice, including the main actions in which the teacher is involved when navigating assessment demands. Lastly, the model discusses the social and emotional dimension of assessment literacy, understanding assessment as a social practice.

Even if the three models conceptualise assessment literacy from different perspectives, they have relevant similarities. First, all of models acknowledge the need for a solid knowledge base as a starting point for developing assessment literacy. Second, all of the models conceptualise assessment literacy as more than a mere set of theoretical contents. In the models presented, teachers need to deal with their own circumstances (e.g. conceptions, attitudes and beliefs); the environment where the assessment will be performed (e.g. tensions, resources, or policies); and the agents involved during the assessment process (e.g. socio-emotional perspective).

To foster teachers' assessment literacy is not the only step for achieving high-quality assessment in universities, but it is a necessary one. Its effects on assessment practices and teachers' well-being have been studied (Popham, 2009), and calls are being made to include assessment literacy enhancement programmes throughout teachers' professional life (Xu & Brown, 2017). One way of achieving assessment-literate teachers is via formal teacher education, which is another popular field in educational research.

Teacher Education

Research on Teacher Training Programmes in Higher Education

In recent years, several studies have contributed to generating knowledge about how such training programmes work, and how effective they are in improving the quality of teaching and learning. Despite the existence of decades of studies on this topic (e.g.

Ferguson & Womack, 1993; Guyton & Farokhi, 1987), according to Gibbs and Coffey (2004) confidence in teacher training programmes is not yet based on empirical results (Coffey & Gibbs 2000; Norton et al., 2005). The literature about the efficacy of teacher training programmes presents mixed results. On the one hand, Gibbs and Coffey (2004) studied the impact of teacher training programmes on teachers' pedagogical skills and their conceptions about teaching and learning. In their study, the results showed that, after the training programmes (lasting between 4 and 18 months), the teachers participating in the course changed the focus of their instruction from one centred on the teacher to one more focused on the student. In addition, the teacher's pedagogical skills improved significantly, according to the opinion of the students. The teacher training programme also had effects on the students themselves, who moved from a superficial approach to learning (following the terms of Ramsden, 1991) to a deeper one.

On the other hand, Norton et al. (2005) have stated that we still do not have enough evidence to affirm with certainty that training programmes have a significant impact on teaching practice. In their study, they found no differences between the control group and the teachers subject to training programmes, as is also the case in studies such as Ödalen et al. (2019). Postareff et al. (2007), on the contrary, did find significant differences in terms of approaches to teaching after training programmes. Furthermore, when these programmes reach a certain duration (30 ECTS), they also found a notable increase in teachers' self-efficacy.

In any case, despite the existence of literature on the operation and effectiveness of teacher training programmes in higher education, it is evident that there is still no consensus on the matter, which thus requires more research in the field. It is noteworthy that most of the existing literature deals with the effect of training programmes on approaches to teaching and learning at a general level. There are even fewer studies

focused on evaluating the effectiveness of these programmes in terms of specific aspects of learning, such as assessment.

In-Service Training Programmes on Assessment in Higher Education

Several authors detail the needs for teacher education, even for teachers already in practice. In-service teacher education programmes, often provided by universities, can serve as a powerful way of promoting assessment literacy among teachers – if done properly (Gibbs & Coffey, 2004). The advantage of in-service training is the possibility for teachers to experience in real time the new concepts and ideas discussed during training, either through their own practice or in community with other professionals. Several authors have discussed the value of the role of communities of practice, networking and peer interaction in supporting professional learning (Simons & Ruijters, 2004; van Schalkwyk et al., 2012; Xu & Brown, 2016) and the importance of situated learning in facilitating transfer (Lave & Wenger, 1991).

Many authors offer recommendations for designing an effective teacher training involving assessment. DeLuca (2012) highlights the importance of assessment education taking various forms and integrating different stakeholders' perspectives. As has been previously stated, assessment design and implementation are complex processes (Bearman et al., 2017) that involve many variables and perspectives. Accordingly, assessment training must reflect this complexity by integrating heterogeneous views. In the next paragraphs we expand on recommendations summarized by Xu & Brown (2016) for teacher training on assessment.

First, Sato et al. (2008) ask for assessment literacy to become a part of teacher accreditation and certification. This would not be the case among higher education teachers in several countries, but it does open an opportunity for designing formal training on assessment for in-service teachers. This would entail a proper evaluation of teachers'

assessment literacy, and the requirement to perform activities designed for their training to earn a certificate.

Second, Graham (2005) highlights the importance for mentors to pay attention to teachers' prior beliefs on assessment. This would require individualised attention by the instructor, which is only achievable with a limited student teacher–instructor ratio, and with programmes of reasonable length. However, as assessment literacy is a phenomenon involving many teachers' individual variables (e.g. conceptions about assessment), addressing these variables through individualised attention would enormously increase the efficacy of such teacher training programmes.

Lastly, Lam (2015) recommends that the training content be localised, subject-area specific to allow for teachers' free choice. Once again, this would require a considerable quantity of resources to make the assessment training as personalised as possible. However, as well as individual factors, there are several contextual factors that are relevant in terms of assessment design and implementation. As not every teacher shares the same context, the possibility for them to choose a training that fits their individual environment would be highly beneficial.

These recommendations offer a framework for what can be considered good teacher training practices for assessment. As proposed, in-service training could be important in developing assessment literacy among higher education teachers. This is especially critical in Spain, as there is no compulsory pre-service pedagogical training for higher education teachers. However, many universities offer a large set of in-service training courses on a variety of topics. To analyse whether or not these training courses are in line with the theoretical recommendations is the aim for the third study in this thesis.

We have taken a tour of the importance of educational assessment and its worrying current state in higher education. We then explored the role of teachers as assessment agents, the way in which they make decisions and the factors that may influence them in making assessment decisions. Lastly, we explored what knowledge and skills teachers should have regarding assessment, and how teachers should be trained to become assessment literate. These are the key theoretical points which support this doctoral thesis.

Aims and Empirical Planning

Higher education is a particularly relevant educational level for students, given its role as the last formal preparation before the beginning of professional life. As such, and given its accessibility, there are numerous studies exploring this educational level from a large number of approaches focused on both the student and the teacher. One of the most studied areas in this level is the educational assessment. Having high-quality assessment is especially relevant in higher education, because university students are supposed to be independent and capable of a greater extent of self-regulation. This, added to the difficult context in many classrooms, makes interactions with teachers less frequent, and the proper design of these interactions is thus more important.

Despite this, the literature on assessment in higher education has two important gaps that constitute the aims of this doctoral thesis. The first of these is the figure of the teacher as assessor, exploring the teacher's role before implementing assessment methods and the decision-making carried out during the process. To support this aim, two studies are proposed. The first study is focused on the contextual factors related to assessment design – especially disciplinary differences. In this study, we compared assessment design and implementation in three programmes (sport science, mathematics and medicine) at four Spanish universities. Using a mixed-methods approach, data from the syllabuses and semi-structured interviews with teachers were analysed.

The second study is focused on personal factors related to assessment design. This study aims to identify different profiles of higher education teachers based on the way they design their assessment methods. It also explores differential characteristics in each profile and the differences in the assessment methods preferred by the teachers and those implemented in their subjects. To do so, 16 teachers from four universities participated in a think-aloud simulation task: designing the assessment methods for a set of learning outcomes.

The second research gap is the study of teacher training programmes and their effect on assessment practices. The results of the two previous studies are aimed at facilitating the design of more appropriate and effective training programmes for higher education teachers. Starting from the theoretical models about assessment literacy, the third study presents a nationwide analysis of in-service teacher training courses about assessment. We screened all of the professional development training courses from Spanish public universities (N = 1627). Data from all of the assessment courses (N = 82) were collected, analysed and compared with the scientific work on assessment literacy.

The set of three studies – and therefore this thesis – expands the empirical knowledge about how higher education teachers design and implement their assessment methods through a broad variety of data collection and analysis methods. This aims to create an impetus towards the design of assessment methodologies more consistent with scientific knowledge, which will ultimately lead to an improvement in educational quality in higher education programmes.

CHAPTER 3

Assessment in higher education across different disciplines: Design and Implementation in Three Undergraduate Programmes

Javier Fernández Ruiz[1], Ernesto Panadero[2, 3] & Daniel García Pérez[4]

Author Note

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

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

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

[4] Departamento de Psicología. Universidad Europea de Madrid

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

Abstract

The role of the academic discipline is a major factor in the assessment design and implementation in higher education. Unfortunately, a clear understanding of how teachers from different disciplines approach assessment is still missing; this information can lead to teacher training programmes that are better designed and more focussed. The present study compared assessment design and implementation in three programmes (sport science, mathematics, and medicine) each representing a discipline from 4 Spanish universities. Using a mixed-methods approach, data from syllabi (N= 385) and semi-structured interviews with teachers (N=19) were analysed. The results showed different approaches to assessment design and implementation in each programme in two main axes: summative or formative purposes of assessment, and content-based or authentic assessment. Implications for further research are discussed.

Keywords: Assessment design, assessment methods, physical education, mathematics education, medicine education

Assessment From a Disciplinary Approach: Design and Implementation in Three

Undergraduate Degrees

Historically, assessment has been one of the main areas of study in educational research (Baird et al., 2017). Importantly, in the last decades, scholars have focused on the formative uses of assessment to improve learning and instruction (Wiliam, 2011) with a special focus on higher education (Boud & Falchikov, 2007). University students are supposed to be more autonomous and thus they have fewer opportunities for direct interaction with the teacher (Brinkworth et al., 2009). Therefore, it is especially important to guarantee that the university assessment practices counteract those fewer interactions – e.g., via frequent feedback – to ensure students experience learning benefits. With such purpose, several authors have explored the type of assessment practices usually implemented in higher education and their implications in terms of learning (e.g., Ibarra-Sáiz & Rodríguez-Gómez, 2010; Lipnevich et al., 2020, 2021). From this line of work, we know that different assessment methodologies lead to different learning outcomes and that they vary among disciplines (e.g., Authors, 2019) – thus, the need to investigate the disciplinary differences.

Assessment practices in higher education are sometimes far from what can be considered formative assessment (Wu & Jessop, 2018) and particularly, studies such as Authors (2019) and Jessop and Tomas (2017) show a traditional assessment scenario in European universities. This is not necessarily an effect of a lack of effort from university teachers who spend a third of their time designing and promoting assessment activities (Izci & Caliskan, 2017). It actually has been argued that it is due to a lack of training and knowledge about assessment techniques that university teachers do not reach the full potential of their assessment in increasing their students' learning (Stiggins, 2007; Authors, 2021).

To understand university assessment practices, it is crucial to explore the conditions under which lecturers design assessment. Importantly, university teachers should have different levels of control regarding their instructional methods (i.e., academic freedom principle). In this vein, teachers usually have freedom to choose specific assessment practices unless they are stablished

at the institutional level –e.g., the department (Authors, 2019). As an example, Spanish universities offer general guidelines such as the number of times that a student can fail a subject, or the need to implement at least one assessment task complementary to the final exam (Gómez et al., 2013), but the rest of assessment decisions are usually made by the teachers. Additionally, we need to understand the specific processes by which university teachers make their assessment decisions. Indeed, the methods used and the factors that influence that assessment design process require further study (Bearman et al., 2017). Therefore, our aim is to investigate how university teachers from three disciplines design their assessment practices.

Assessment Design by University Teachers

We define assessment design as "all processes that take place in order to form specific assessment tasks for a particular course or unit, including the selection and timing of tasks, development of rubrics and redevelopment of a task in response to student performance" (Bearman et al., 2017, p. 50). In our study, we focus on the selection of assessment tasks and the importance given by the teachers to each one, represented by its weight on the students' final grade.

Unfortunately, there is little information regarding the assessment design process, with only a few studies published in recent decades. Studies from McMillan (2003) and Meyer et al. (2010) found a tension between several personal and contextual factors that affect assessment design processes and outcomes. Among the personal factors, a key factor analysed is the relationship between assessment conceptions and practices. Postareff et al. (2012) found that most university teachers have difficulties describing the purpose of their assessment practices. Other studies showed a link between teachers' conceptions about assessment and the assessment methods implemented (Brown et al., 2009; Fernández Ruiz & Panadero, 2020) or the students' involvement in assessment (Panadero et al., 2014). Regarding the contextual factors, the influence of policies and regulations, departmental cultures, available resources, and the discipline are usually listed (Meyer et al., 2010). The interlinked nature of the personal and contextual factors is acknowledged in previous works (Bearman et al., 2017) and would lead to unique individual

challenges where some teachers experience dissonance between the assessment they want to implement and the one they are actually using.

Crucially, in recent years, there has been a development with a few scholars exploring the assessment design process in detail in contrast with the previous literature that was mainly theoretical. First was the above-mentioned work by Bearman and colleagues (2017). Secondly, Authors (2021), using a think-aloud method, found three different profiles to assessment design. These approaches were: (a) *classic*, when teachers focused on the feasibility of the assessment methods proposed; (b) *competence*, when teachers focused in the alignment between learning outcomes and assessment methods; and (c) *cohesive*, when teachers focused on the alignment between teaching and assessment methods.

Importantly, our 2021 publication and the present article are founded in the Assessment Design Decisions Framework (see Table 3.1), developed by Bearman et al. (2016), mainly for two reasons. First, this framework is the most complete classification available covering the most relevant variables in the design of assessment practices. Second, the framework was conceived as a deductive model starting from empiric evidence, and it was designed after several interviews with higher education teachers from different disciplines. These two characteristics make the Assessment Design Decisions Framework the most adequate framework for analysing our data.

Table 3.1

Assessment design decisions framework. Bearman et al. (2016) p. 552

Purposes of assessment
How can assessment: (1) support student learning; (2) generate grades that will form part of subsequent certification; and (3) equip learners for making future judgements?
Contexts of assessment
Which of the following attributes needs to be considered in assessment design? What specifically about each can be taken into account? How can tensions between different needs be reconciled?
• Characteristics of learners/students
• Institutional assessment principles and policies
• Professional, vocational or employment-related requirements
• Departmental, disciplinary and personal norms, expectations and ideas
• The overall program and the role of the course/module
• Learning environment, for example, mode (online/face-to-face/blended); class size
Learner outcomes
How does assessment align with, and promote, desired learner outcomes, including:
• Course/module learning outcomes.
• Overall program learning outcomes.
• Professional requirements
• Learners' general professional or intellectual development.
Tasks
• What is the rationale for each task?
• How do the tasks drive learning? What do the tasks specifically require learners to do?
• How will successful completion be judged?
• How are tasks best distributed across the semester?
• How will students contribute?
• Which tasks will be graded?
Feedback processes
• How are multiple feedback opportunities achieved through the distribution and relationship of tasks across the course/module/overall program?
• What types of feedback information will be provided and by whom?
• How will learner performance be used to influence the (re)design of later tasks?
Interactions
• How will resistance or engagement from learners or colleagues influence assessment processes?
• How will learners understand what is required in the assessment task(s)?
• What information will be needed to improve this assessment for subsequent occasions?
• What associated changes in teaching and learning activities will be required?

Starting from the Assessment Design Decisions Framework, we focus our analysis on understanding how teachers define the purpose of assessment, which contextual factors they consider when designing their assessment methods, and how assessment is aligned with learning outcomes. Specifically, we explore those areas from a disciplinary lens.

The Influence of Disciplines in Assessment Design

Disciplinary signature practices are a rising topic among instructional science literature (see, for example, a recent special issue in this journal; Quinlan & Pitt, 2021). Some studies have analysed the disciplinary effect of the assessment methodologies used in higher education (Lipnevich et al., 2020, 2021; Authors, 2019). Accordingly, differences among disciplines regarding the assessment methods implemented have been found in countries such as the US (Lipnevich et al., 2020) and Spain (Authors, 2019). However, as Quinlan and Pitt (2021) argue, most of the assessment related literature still tends to focus on generic concerns, with little attention to the specific disciplinary contexts (Esterhazy, 2018; Wiliam, 2019). Several authors argue that assessment activities require disciplinary interpretation to be fully understood (Bearman et al., 2016, 2017; Boud et al., 2010). That is why the study of the similarities and differences among disciplines can help to offer specific recommendations for greater adjustment of the instructional context. This endeavour would ultimately increase the effectiveness of assessment and facilitate its implementation.

Importantly, there are only a few studies that have compared assessment practices among disciplines. Among them are the above mentioned by Authors (2019, 2021). Additionally, Carless (2015) analysed the assessment implementation in three disciplines from the University of Hong Kong: history, architecture, and law. He found substantial differences in the assessment methods and feedback delivered. Nevertheless, assessment design was not the main topic of that study, and the sample included only award-winning teachers, making it difficult to be generalised. Bearman et al. (2017) interviewed 33 Australian teachers about their assessment design and found the influence of disciplinary traditions as a barrier for change or innovation regarding assessment practices. However, as the authors state, "the role of disciplinary approaches may be significant and remains an area for future research" (Bearman et al., 2017, p. 49). For these reasons, here we compared assessment design in three disciplines. We will next expand on the rationale for the selection and the differences found.

Selection of Disciplines and Specific Degrees

Our selection of disciplines was based on previous studies that found clear differences in the assessment patterns among those three disciplines (Authors, 2019, 2021). Authors (2019) conducted a study comparing the assessment methodologies implemented among all disciplines using a nationwide sample of syllabi. Salient differences were found among all disciplines, but here we present the three that were chosen for this study. First, we found that education-related disciplines, such as sport science, showed a greater use of portfolios, group assignments, and attendance. This approach can be interpreted as the intent of continuous assessment. Second, health science-related disciplines, such as medicine, showed a greater use of practical examinations, which is consistent with the new paradigm described in the literature. Lastly, third, hard science-related disciplines, such as mathematics, showed a greater use of partial examinations and, to a lesser extent, practical examinations, aligned with a more theoretical approach. Also, Lipnevich et al. (2020, 2021), conducting a similar study in the United States, found greater use of final examinations in science-related disciplines.

To choose a specific programme as representative of each discipline, some of the most salient ones were also selected considering their availability. The three degrees selected were: sport science for education, mathematics for hard science, and medicine for health sciences. These degrees are widely offered in Spanish universities. The number of disciplines analysed was subrogated to the type data collection as this is an in-depth, qualitative study. To understand each disciplinary context and minimise the bias of each teacher's particular views and experiences, the authors agreed that each discipline had to be represented by at least four teachers.

Importantly, to have a stronger rationale for the selection, we also explored the limited existent evidence in those disciplines. Sport science teachers have great interest in innovative pedagogies. Studies, like those by Barba-Martín et al. (2020), Lopez-Pastor et al. (2013), and Hay (2006), detailed various approaches to alternative forms of assessment, that seem to be common among teachers in this discipline. However, traditional assessment practices still have a significant space, especially regarding the lack of involvement of students in assessment. For example, Fraile et al. (2018) carried out a census study of the syllabi of a sport science degree in

Spanish universities finding that only 7.55% of teachers implemented self-assessment. This proportion is in line with other disciplines, as has been reported in other studies (e.g., Panadero et al., 2014).

The assessment practices in medicine have undergone a remarkable change since the 1990s (Govaerts, 2008). This change has transformed the assessment from a theoretical and content-based approach to one aimed at simulating professional practice as closely as possible. In this new paradigm, two methodologies stand out (Schuwirth & Van der Vleuten, 2019). The first one is simulation in real or virtual environments the medical praxis. The second is work-based assessment, exemplified in this case by a large amount of clinical work assessed with a practical approach. Both assessment methodologies based in authentic assessment principles emphasise the interest to prepare students for their future practice (Cumming & Maxwell, 1999).

In the case of mathematics, the literature on the specifics of their assessment methodologies is scarce (Howard et al., 2019). However, Trenholm and colleagues (2015) highlighted the summative and theoretical approach that characterises assessment in this discipline. Iannone and Simpson (2011) reported eight assessment practices common in mathematics: closed-book examination, dissertation, open-book examination, multiple-choice test, oral examination, regular example sheets, and project presentation. While these methodologies may seem different, they share an emphasis in the summative assessment of results above a formative assessment of the processes, an approach that has been mentioned in other studies (e.g., Lipnevich et al., 2020).

The National Context

To understand the context of this the study, is it necessary to address the influence of the Bologna process. The present study is based in Spain, one of the countries participating in the Bologna process. This process aimed to ensure standards and quality of higher education credentials across the European Union countries (Watcher, 2004), with a special focus on continuous education. As Haukland (2020) argues, even if the Bologna Process has made a strong

impact on European higher education, its greatest impact has come with the national reforms introduced to comply with the process.

In Spain, these reforms by the Bologna Process propelled a change in teaching/learning methodologies. The focus went from teaching to learning, from academic exposition to students' involvement, and from teaching content to teaching how to learn (Gil & Beltrán, 2018). After its implementation, in many European universities, teachers had to redesign their teaching and assessment methods to adapt them to the new plan, making this moment an ideal context to explore how these redesigns were carried out. Since then, Quesada-Serra, Rodríguez-Gómez, and Ibarra-Sáiz (2017) have highlighted the concern of Spanish teachers for carrying out assessment-related innovations. Based on the literature, there has been an improvement over the last two decades in the Spanish assessment panorama: teachers now use a greater variety and number of instruments to assess their students' learning, but the prevalent assessment profiles are still traditional, with barely any student involvement in assessment (Fraile et al., 2018; Panadero et al., 2014, 2017).

Research Aim and Questions

This study aims to explore the similarities and differences in how teachers in three disciplines design their assessment methods. We explored three main research questions:

RQ1 – What assessment methodologies are implemented in each discipline?

RQ2 – What adjustments have been made by teachers in recent years?

RQ3 – How do teachers in each discipline justify the assessment methodologies they implement?

- o *RQ3.1 What is the main purpose of assessment?*
- o *RQ3.2 What factors determine the learning outcomes?*
- o *RQ3.3 What external factors determine the assessment methodologies?*

Method

Participants

Our study involved three disciplines and four different Spanish universities. Eighteen interviews with 19 teachers were performed (one interview included two teachers from the same course). This study used the mixed methods approach to sampling, as preliminary quantitative analysis was performed to select suitable and representative participants via stratified sampling (Teddlie & Yu, 2007). Sixteen participants were selected by stratified sampling according to the data in their syllabi, assessment profile variables (traditional versus innovative), teaching experience (novel versus experienced), and gender. The remaining three participants were chosen by availability. However, their assessment profile and demographics (teaching experience and gender) were also analysed using the same inclusion criteria as the rest to ensure that they kept the sample balanced in the variables mentioned. Participants' demographics are displayed in Table 3.2.

Table 3.2

Participants' demographic data.

	Gender	University	Faculty	Teaching experience (years)	Current unit experience (years)
1	Male	1	Sport Sciences	31	16
2	Male	1	Sport Sciences	20	8
3	Male	1	Mathematics	20	3
4	Female	4	Sport Sciences	6	4
5	Male	2	Mathematics	31	1
6	Teaching pair (male + female)	1	Medicine	15	3
7	Male	1	Mathematics	28	2
8	Male	3	Sport Sciences	17	13
9	Female	2	Medicine	15	5
10	Male	1	Mathematics	28	1
11	Female	4	Sport Sciences	10	7
12	Male	2	Medicine	38	25
13	Female	1	Mathematics	6	2
14	Male	2	Mathematics	20	6
15	Male	4	Sport Sciences	14	11
16	Male	1	Sport Sciences	20	15
17	Male	2	Medicine	44	42
18	Male	4	Sport Sciences	10	7

Materials

Syllabi

A syllabus is an official, mandatory document that contains information about the teaching and assessment methods. Some scholars have stated that syllabi provide an interesting view into the instructional environment teachers create in their university courses because it is an unobtrusive but powerful indicator of what takes place in classrooms (Bers, Davis, & Tylor, 2000). The syllabi included every course from the selected universities. All the universities had similar characteristics (public and large) but as not every discipline was available in each university, the distribution is unequal. University 1 includes the three disciplines, but university 2 only includes mathematics and medicine, so additional data about sport science was collected from universities 3 and 4. Full data about distribution are presented in Table 3.3.

Table 3.3

Sa*mple distribution in programme and university.*

	Sport sciences	Mathematics	Medicine
Uni 1	57 *(3)*	39 *(4)*	81 *(3)*
Uni 2	.	65 *(2)*	45 *(2)*
Uni 3	97 *(2)*	.	.
Uni 4	0 *(3)*	.	.

Note: Syllabi analysed *(lecturers interviewed)*

Semi-structured interviews. The interviews included 13 questions, based on those designed by Bearman et al. (2017). After personal communication with those authors, some modifications were made to the original prompt. Other minor changes were made to the interview prompt after the first interviews to further explore some topics mentioned by the participants. The interviews were conducted in the participants' native language (Spanish) and the average length was 45 minutes. A complete interview prompt can be found in Appendix A.

Data Collection Methods

We employed a sequential mixed-methods design (Creswell & Creswell, 2017). It included two main data sources: course syllabi and semi-structured interviews of the teachers. Quantitative data from the syllabi were used to select the interviewees and as a tool to prepare for and interpret the interviews. Syllabi data was downloaded from the university's webpage. From each syllabus we extracted the information on assessment practices and the percentage of each on the students' final grade (more information in Appendix B).

Procedure

Syllabi from university 1 were collected and analysed between November and December 2018. As a result of this analysis, 24 teachers from university 1 were invited for an interview; however, only 10 agreed to participate. The first author conducted three preliminary interviews as a pilot, to assure that the questions were adequate and understandable. These interviews are not included in our final sample. To enlarge the sample, we repeated the process of collection and analysis of syllabi, this time from universities 2 and 3. Another six interviews were performed with teachers from these universities. Lastly, three additional teachers from university 4 were contacted (selected by proximity to the authors) and agreed to participate. Interviews took place from February to November 2019.

Data Analysis

Both quantitative and qualitative descriptive approaches (Sandelowski, 2010) were used depending on the data source. Data from the syllabi was used to answer RQ1. Regarding the syllabi, descriptive analysis included means and standard deviations ($N = 385$). Data from the interviews was used to answer RQ 2 and RQ3. Following the open design of qualitative studies (Creswell & Creswell, 2017), six interviews were performed and analysed before conducting the rest of the interviews. We used theoretical coding, based on the framework proposed by Bearman et al. (2017). In two moments (after six interviews and at the end of data collection), the first and third authors triangulated the codes and the conceptual analysis through several online

discussions. After each discussion, once consensus was reached, the first author cleared the code list and edited the final database. Due to the interpretative nature of qualitative data, an open discussion between the authors was preferred over a quantitative interrater analysis (Bazeley, 2013; Creswell & Creswell, 2017).

The data from the interviews is represented as handmade figures in the results section. In those figures, the size of each circle represents the popularity of this topic among the participants' responses, with the numbers inside indicating the percentage of teachers from each discipline mentioning the concept. The position of each circle in the hexagon represents the differences in popularity: the closer to one of the disciplines, the more prevalence it has, while a concept mentioned equally among the three disciplines would be placed in the centre of the hexagon.

Results

RQ1 – What Assessment Methodologies Are Implemented in Each Discipline?

To explore this research question, we explored the syllabi data. As shown in Figure 3.1, the assessment methods reported in the syllabi showed relevant differences among disciplines. Sport science syllabi are characterised by fewer theoretical examinations, which is balanced by more continuous assessment methods, especially assignments. In mathematics, the situation can be described as the opposite. Many partial and final exams are carried out, and attendance provides greater weight in the students' final grade. Finally, in medicine, there is less variety in the assessment methods: a final exam is the most popular option.
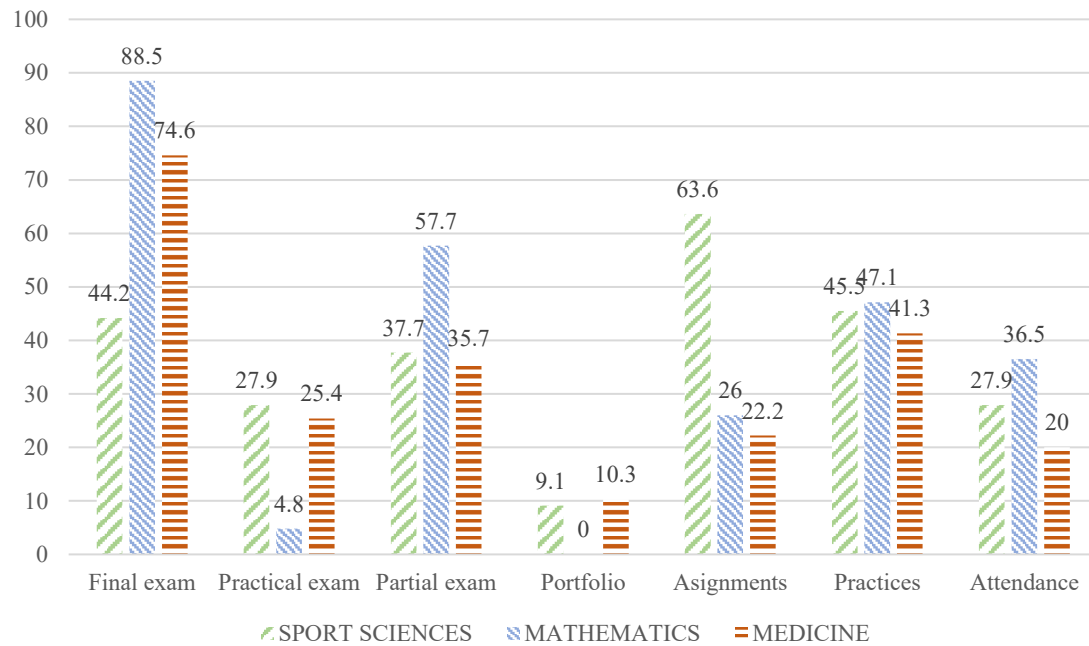
Figure 3.1. Assessment evidences used in each discipline.

RQ2 – What Adjustments Have Been Made by Teachers in Recent Years?

Teachers were asked about the areas of interest regarding their development of assessment methodologies. As Figure 3.2 shows, there are clear patterns in sport science and mathematics. Sport science teachers focused on improving the quality of the feedback offered to their students (63%), either by increasing its quantity or by incorporating feedback methods other than grades, such as rubrics. In addition, there was an interest in promoting student participation in assessment via self or peer assessment.
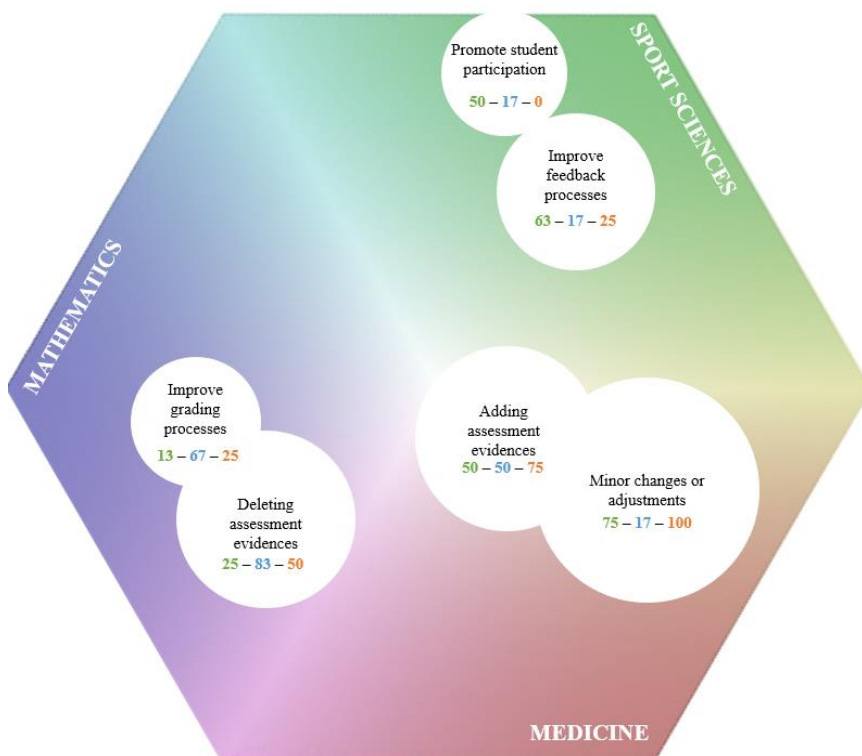
Note: Numbers represent percentage of lecturers mentioning this category
in Sport sciences – Mathematics – Medicine.

Figure 3.2. Areas of interest on the assessment methodologies development.

In mathematics, teachers have made changes to their assessment methods, which are aimed primarily at improving grading processes (67%). The most frequent way of achieving this was to modify the weighting of the different instruments in the final grade. This had two different aims: to represent adequately the student's effort, and to force students to work harder in the tasks that the teachers consider more important. Furthermore, preventing students from cheating was also a common concern among teachers of this programme.

I had to change my assessment methods because my students were cheating all the time. Even the students themselves, those who actually worked hard, complained about it.

Mark – mathematics

Note: All the interviewees are represented in the manuscript with aliases.

There was not a clear pattern in medicine. Similarly, to those from other disciplines, teachers reported an effort to incorporate new assessment evidence to their subjects. Teachers

from medicine are also more likely to implement minor changes or adjustments to their assessment methods (100%), mostly regarding its format or its periodicity. These changes use to be the result of students' response each year.

RQ3 – How Do Teachers in Each Discipline Justify the Assessment Methodologies They Implement?

Following the framework proposed by Bearman et al. (2016), we divide this research question in three different areas: purpose of assessment, learning outcomes, and external influences.

RQ3.1. What is the main purpose of assessment? In this case, medicine shows the clearest tendencies as shown in Figure 3.3.
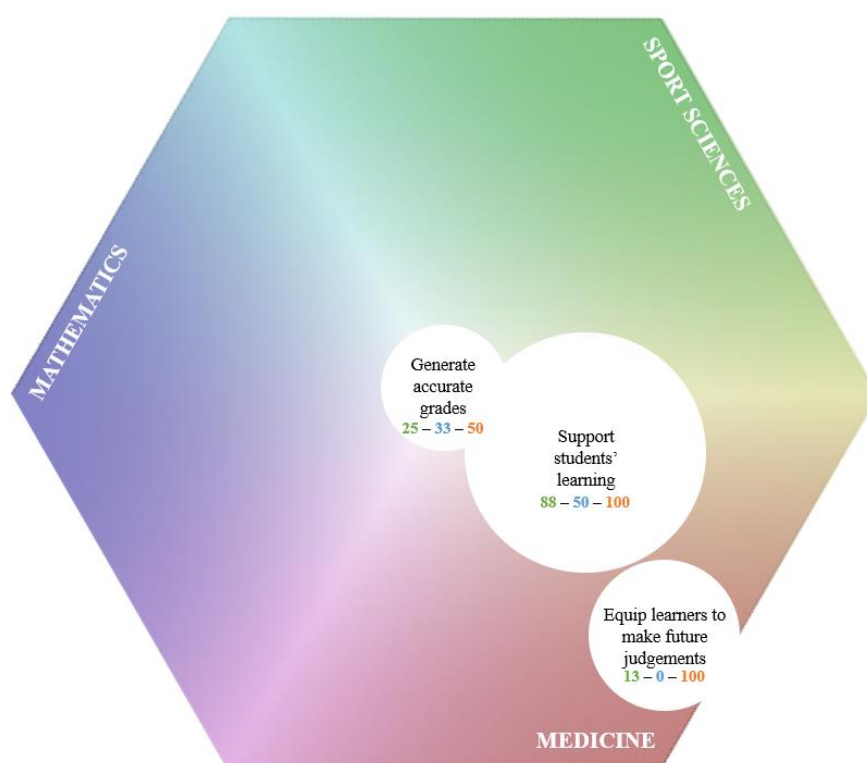
Note: Numbers represent percentage of lecturers mentioning this category
in Sport sciences – Mathematics – Medicine.

Figure 3.3. Purpose of assessment.

Teachers' answers mostly differed in the formative vision of assessment. Sport science (88%) and medical teachers (100%) were more inclined to understand assessment as a tool for

supporting and improving student learning. In addition, all medical teachers mentioned the use of assessment to equip their students with the ability to make better judgments. Teachers' responses showed an important concern about the need for critical thinking in their students' future profession and an attempt to provide them with this skill through assessment. Lastly, teachers from the three disciplines equally agreed that an assessment must generate accurate and fair grades.

RQ3.2. What factors determine the learning outcomes? Teachers described two main factors that determine the learning outcomes that are assessed: (a) the professional context in which the students must perform and (b) the content of the unit considering its role in the entire degree. Importantly, there were remarkable differences between these two among the disciplines (Figure 3.4). In medicine, the professional context is usually very specific, being easier for teachers to prepare their students for their future job.

> In medicine what you study is exactly what the patients will have in the future. In other words, there is no discrepancy between the theoretical element and the professional reality.
>
> *Victor – Medicine*
>
> **Note**: All the interviewees are represented in the manuscript with aliases.
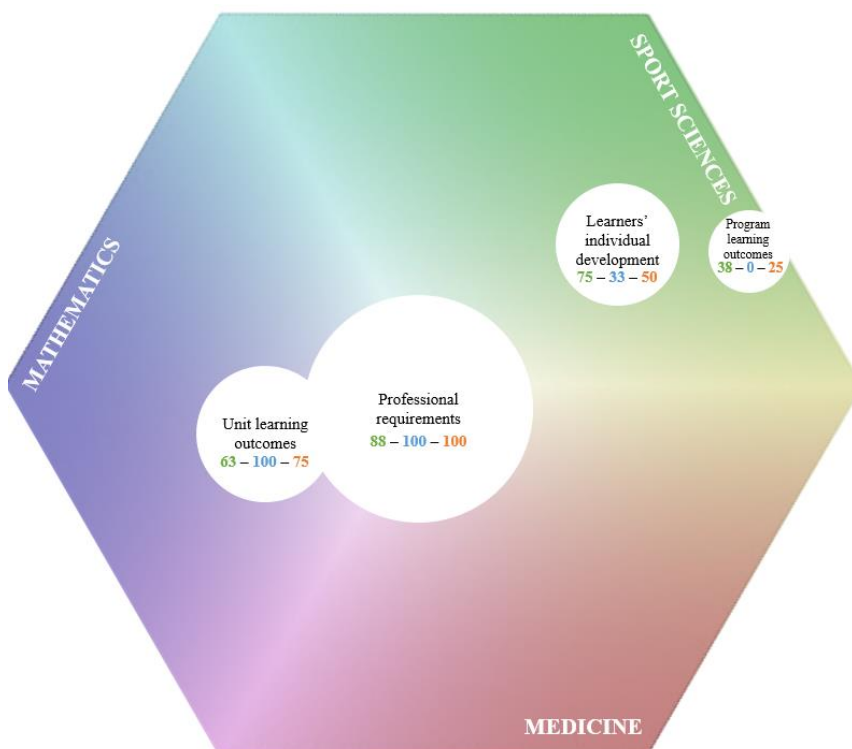
Note: Numbers represent percentage of lecturers mentioning this category
in Sport sciences – Mathematics – Medicine.

Figure 3.4. Factors determining the learning outcomes.

This approach did not happen in mathematics, where career opportunities are more diffused. For this reason, the mathematics teachers seemed unsure of the job requirements that their students will face. As all of them mentioned professional requirements as a factor, it was strictly to link them with the specific content of their subject.

It is difficult to answer, because in mathematics there are lots of professional possibilities. Depending on what you do, it will be related with my assessment methods or not... But there are people doing very different things, and I cannot guarantee that all I am trying to assess will be applied in everybody's future jobs.

Mark – mathematics

Note: All the interviewees are represented in the manuscript with aliases.

Even if professional requirements (88%) and unit learning outcomes (63%) are fairly popular topics among sport science teachers, they are less inclined to consider them in comparison

with the other two disciplines. Instead, they tend to focus on the learners' individual development (75%), adopting a competence approach that is not necessarily linked with their students' employability.

RQ3.3. What external factors determine the assessment methodologies? As shown in Figure 3.5, "departmental requirements" and "learning environment and infrastructure" were the greatest determinants in assessment design.
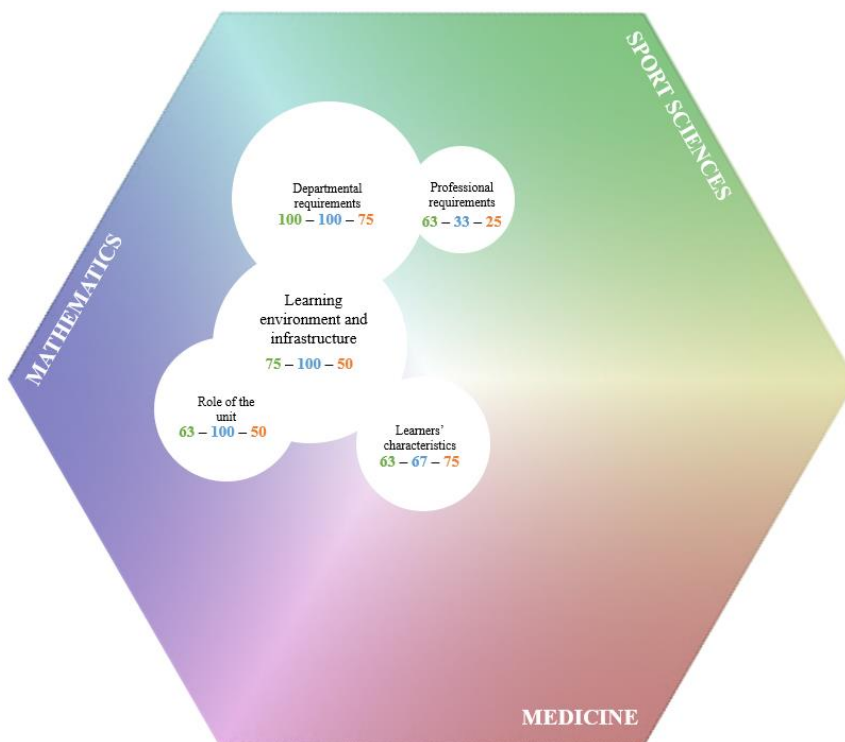
Note: Numbers represent percentage of lecturers mentioning this category
in Sport sciences – Mathematics – Medicine.

Figure 3.5. External factors determining the assessment methodologies.

Teachers report that the number of students is a major factor to consider when they are designing assessment methodologies.

> You use exams because of the number of students. The ratio can make you feel pressured. If I had one hundred students . . . it would be impossible for me to keep using this assessment method.
>
> *Charles – sport science*

However, the student–teacher ratio was not the only factor related to the available resources that influence assessment design. Other factors such as the available technology, size and distribution of the classrooms, and the availability of specific materials were also frequently mentioned as a limitation.

The characteristics of learners were mentioned by teachers in the three disciplines, albeit for different reasons. In sport sciences, teachers report great variability between different cohorts, which makes them adapt their methods to their current students. In medicine, teachers reported how the students were in favour of traditional assessment methods, such as multiple-choice exams, so that they would be prepared for the national examination board (MIR) that they have to pass in order to become doctors in the public health system. In mathematics, most of the teachers referred to the increase in the overall quality of the cohorts, and how it affects their practices.

In mathematics, the mean level has raised spectacularly in the last five years… This course I taught to first-year students and only one failed. One out of 60 students. I assure you that this is not normal in mathematics. In the end, consciously or not… you end up raising the level.

Edward – mathematics

In sum, the discipline seems to have effects on the way teachers justify the design of their assessment methods. Sport science teachers showed a greater interest in the formative component of the assessment and linked it to the personal development of their students. In medicine, teachers showed interest in developing critical judgement in their students, but they also think about their students' professional future. In the case of mathematics, teachers tended to focus on the summative assessment of the contents of each course and on the internal coherence of the degree itself.

Discussion

This study aimed to explore the similarities and differences in how teachers in three disciplines design their assessment methods. To do so, this study explored three research questions whose results will be discussed next.

Our first research question explored which assessment methodologies are implemented in each discipline, and the second research question explored what adjustments have been made by the teachers in the recent years. For readability issues, they will be discussed together. The three disciplines show distinct profiles regarding the design of assessment methods especially in whether the teachers used formative or summative approaches.

Sport science teachers showed a marked formative approach in their assessment design, a possible effect of their training in good pedagogical practices. As reported by López Pastor et al. (2013), teachers from this discipline also emphasise putting innovative methodologies into practice. In our results, this is translated into a more continuous assessment method, limiting the presence of final examinations, and substituting them for assignments during the course. Teachers from this discipline were also more concerned about feedback quality and students' participation in assessment. This approach might be linked with their previous training, as several of the teachers interviewed were trained as secondary teachers in physical education, and others had bachelor's degrees in teacher education or pedagogy. This previous training would explain their interest in making their assessment formative and beneficial for their students' learning.

In mathematics, the situation was the opposite of sport science. Teachers are usually focussed on the summative aspect of their assessment methods, as has been discussed in previous studies (e.g., Trenholm et al., 2015). Closed and open book examinations are widely used, in line with previous findings (Simpson, 2011), but final examinations have lost weight in favour of partial examinations after the Bologna regulations. Another sign of this summative trend is how mathematics teachers make modifications to their assessment methods almost exclusively related

with grading. Most of the interviewed teachers are also sceptical about the implementation of non-graded tasks or students' participation in assessment.

Teachers from medicine show a mixed approach to the formative-summative tension. They acknowledge the importance of grading, and final exams are present in almost all the courses, probably with the aim of training their students for the national examination board. On the other hand, they show a constant effort to incorporate new assessment practices into their subjects, mostly aiming to simulate professional practice (e.g., authentic assessment).

This duality among medicine teachers represents a great example of the difficulties in implementing formative assessment in competitive contexts. In a recent review, Yan et al. (2021) explored the factors influencing teachers' intentions and implementations regarding formative assessment. One of these factors is the "cultural norm," or the societal perception of assessment. This perception can make the implementation of formative assessment practices extremely challenging in places with great examination culture, as it is the case of many Asian countries such as China (Brown & Gao, 2015; Yan & Brown, 2021). On a smaller scale, this could be also the case in medical degrees, considered one of the most academically demanding training programmes out of any profession (Tian-Ci Queck et al., 2019).

Our third research question explored how teachers in each discipline justify the implemented assessment methodologies. Our participants showed a complex system with a great variety of internal and external factors that influence the assessment design. Several of these factors have already been identified in previous studies, such as the tension between summative and formative purposes of assessment (Meyer et al., 2010), the constraints of administrative requirements (Meyer et al., 2010; Norton et al., 2005), and the students' engagement with the proposed methodologies (Watkins et al., 2005). Our results also highlight the influence of the professional expectations of each programme in shaping assessment design. This was not a central topic in Bearman et al.'s (2016) framework, and accordingly, was not considered in the initial design of this study. However, its frequency and relevance in participants' testimony must be addressed in a separate section.

The Influence of Professional Expectations

The differences found in the disciplines analysed show how the disciplinary and professional context can shape the assessment methodologies even more than the characteristics of the degree itself, as other authors have previously stated (Bearman et al., 2017; Carless, 2015). This was especially striking in the comparison between medicine and mathematics.

The students in the former face a very homogeneous professional future, and that makes it easy for teachers to know the skills and knowledge that their students will need for their future employment. This awareness is also because many of the teachers combine their work in teaching with other employment outside the university, which creates an "apprenticeship" approach between them and their students (Harman & McDowell, 2011). That might be the reason for our interviewees' emphasis on the preparation of the students for their professional future, also found in previous studies (Govaerts, 2008).

According to the teachers, the homogeneous professional context helps them to shape the assessment design and implement authentic assessment. Authenticity, understood as realism, contextualization, and problematization when teaching and assessing curricular content (Villaroel et al., 2018), was a major aim among teachers in medicine. Their assessment methods simulated different professional practices, such as scientific conferences, role-plays, or clinical practice. The clarity of their students' future professional paths also allows them to define the competencies needed to become a good professional. This is translated into a more competence-based assessment, instead of a contents-driven one. The competencies assessed by these teachers, unlike the other disciplines, are not only focused on conceptual and epistemological dimensions, but also on social, material, and moral dimensions (Quinlan & Pitt, 2021).

The opposite is found in sport sciences and, especially, mathematics, as the professional path for these students is much more diffuse. This can be because it is a new discipline in the case of sport sciences (Jessop & Mackelar, 2016) or because the professional opportunities have changed in the last years, as in the case of mathematics (Silió, 2019). In these two cases, we found

a less competence-based assessment and one that is more limited to the degree's content. Acknowledging the benefits of authentic assessment for students (Sambell et al., 2013; Bloxham, 2015), it is recommended to explore how to enable the implementation of authentic assessment in disciplines with a diffuse professional context, where students can end up performing highly diverse roles.

Several findings of our study aligned with the findings by Bearman et al. (2016, 2017). First, teachers recognise the existence of an "impetus for change" as a starting point for the redesign of their assessment methods, in similar terms as those reported by Bearman et al. (2017). In the Spanish context, the implementation of the Bologna plan can be considered a nationwide "impetus for change." However, it has not been the only one because the interviewed teachers declare they redesign their assessment methods with a certain periodicity influenced by the students' feedback or, in some cases, as part of a mere trial and error process.

It should be noted that the Assessment Design Decisions Framework proposed by Bearman et al. (2016) is consistent with the teachers' responses, and its use as a coding tool in this study has been organic and relatively simple. This suggests that the factors proposed in the framework can reflect regular teaching contexts. However, the teachers' responses are much more diffuse regarding the purpose of the assessment, and the teachers seem to struggle to describe it clearly, in line with previous studies (Postareff et al., 2012).

Lastly, regarding the level of analysis, it is remarkable how the discipline arises as a much bigger factor of influence than the university. Teachers argue that, apart from general university regulations, they design their assessment practices independently from university tendencies. These university regulations are, on most occasions, vague, and refer mostly to deadlines and revision procedures (Gómez, Sáiz, & Jiménez, 2013). The assessment instruments used and their alignment with the learning outcomes are left in the hands of teachers in all the universities explored. Additionally, there were no significant differences regarding university type. Public and private universities may provide different environments for teachers in several countries (Mohammadi & Karupia, 2020; Álvarez-Castillo et al., 2017). However, in terms of assessment

design, all the teachers interviewed show a similar degree of autonomy, independent of their employment in a public or a private university.

Our results help to explain the reasons behind the presence of discipline assessment practices (Quinlan & Pitt, 2021), which can enrich the discourse about assessment and feedback in higher education. They also support Bernstein's (2000) categorisation of inward and outward-facing disciplines, as clear examples of both categories are found among the analysed disciplines. Doing so, we contend, is more likely to challenge and reframe generic models that have been difficult to put into practice (Quinlan, 2016).

Limitations

Our study has limitations. First, the syllabi and interviews came from four universities from the same region. Second, the participants might be more motivated than the average student because we chose them based on their current assessment practices. Third, our study only covers three disciplines; although they are very different, they cannot be taken as representative of the entire higher education scenario. It is probable that teachers from other disciplines follow different patterns while designing and implementing their assessment. Lastly, while it is true that our data comes from different sources (syllabi and interviews), we did not collect information in how assessment is actually implemented. All these things considered, it is important to remember that our study was an in-depth mixed method with a considerable sample size: 385 syllabi and 19 participants.

Conclusions

This study has shown the similarities and differences in how assessment is designed and implemented among three academic disciplines. Our results have revealed three different circumstances in terms of summative or formative implementations of assessment, highlighting influencing factors such as teachers' previous training or disciplinary culture. This study also revealed different approaches to assessment design depending on the clarity of professional expectations, from pure content-based assessment to authentic assessment. A clear professional

path in their disciplines is needed for teachers to implement authentic assessment in their classrooms. The influence of professional expectations must be acknowledged in assessment-related research. Future research in the field is needed to explore whether different professional panoramas in different countries could lead to different assessment practices even in the same discipline.

References

Álvarez Castillo, J. L., Martínez Usarralde, M. J., González González, H., & Buenestado Fernández, M. (2017). El aprendizaje-servicio en la formación del profesorado de las universidades españolas/Service-learning in teacher training in Spanish universities. *Revista española de pedagogía*, 199-217.

Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart?. *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350.

Barba-Martín, R. A., Hortigüela-Alcalá, D., Pérez-Pueyo, Á., & Sánchez-Santamaría, J. (2020). Conceptual Analysis of Influential Factors in the Motivation and Involvement of the University Student towards the Assessment in Physical Education. *Sustainability, 12(21), 8842. doi:10.3390/su12218842*

Bazeley, P. (2013). *Qualitative data analysis. Practical strategies.* London: Sage

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.

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M. & Molloy, E. (2016) Support for assessment practice: developing the Assessment Design Decisions Framework, *Teaching in Higher Education*, (21)5, 545-556, DOI: 10.1080/13562517.2016.1160217

Bernstein, B. (2000). *Pedagogy, symbolic control, and identity*. Rowman & Littlefield Publishers.

Bers, T. H., Davis, B. D., & Taylor, B. (2000). The use of syllabi in assessments: Unobtrusive indicators and tools for faculty development. *Assessment Update, 12*(3), 4-7.

Bloxham, S. (2015). Assessing Assessment: New Developments in Assessment Design, Feedback Practices and Marking in Higher Education. In *A Handbook for Teaching and Learning in Higher Education*, edited by H. Fry, S. Ketteridge, and S. Marshall, 107–122. 4th ed. Abingdon: Routledge.

Boud, D. and Associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education.* Sydney: Australian Learning and Teaching Council.

Brinkworth, R., McCann, B., Matthews, C., & Nordström, K. (2009). First year expectations and experiences: Student and teacher perspectives. *Higher Education*, *58*(2), 157-173.

Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in education: principles, policy & practice*, *16*(3), 347-363.

Brown, G. T., & Gao, L. (2015). Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes. *Cogent Education*, *2*(1), 993836.

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education, 69*(6), 963–976. doi:10.1007/s10734-014-9816-z.

Creswell, J. W., & Creswell, J. D. (2017). *Research design. Qualitative, quantitative and mixed method approaches* (5a ed.). Thousand Oaks, California: Sage.

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

Fernández-Ruiz, J., & Panadero, E. (2020). Comparison between conceptions and assessment practices among secondary education teachers: more differences than similarities (Comparación entre concepciones y prácticas de evaluación en profesores de Educación

Secundaria: más diferencias que semejanzas). *Journal for the Study of Education and Development*, *43*(2), 309-346.

Fernández Ruiz, J., Panadero, E., García-Pérez, D., & Pinedo, L. (2021). Assessment design decisions in practice: Profile identification in approaches to assessment design. *Assessment & Evaluation in Higher Education*, 1-16.

Fraile, J., Pardo, R. y Panadero, E. (2018). Autoevaluación y autocalificación en el Grado en Ciencias de la Actividad Física y del Deporte: Estudio censal de las guías docentes. *Profesorado. Revista de Currículum y Formación de Profesorado, 22*(3), 163-182.

Gil, L. V., & Beltrán, J. C. H. (2018). The Bologna Process Could Be at Stake: Some Thoughts from Spain. *Universal Journal of Educational Research*, *6*(4), 769-774.

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.

Govaerts, M. (2008). Educational competencies or education for professional competence? *Medical Education, 42*(3), 234- 236. https://doi.org/10.1111/j.1365-2923.2007.03001.x PMid:18275410

Harman, K., & McDowell, L. (2011). Assessment talk in design: The multiple purposes of assessment in HE. *Teaching in Higher Education, 16*(1), 41-52.

Haukland, L. H. (2020). The Bologna Process and HEIs Institutional Autonomy. *Athens Journal of Education*, *7*(4), 365-383.

Hay, P. J. 2006. 'Assessment for learning in physical education'. In *The handbook of physical education*, Edited by: Kirk, D., MacDonald, D. and O'Sullivan, M. 312–325. London: Sage.

Howard, E., Meehan, M., & Parnell, A. (2019). Quantifying participation in, and the effectiveness of, remediating assessment in a university mathematics module. *Assessment & Evaluation in Higher Education, 44*(1), 97-110.

Iannone, P., & Simpson, A. (2011). The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications: An International Journal of the IMA, 30*(4), 186-196.

Ibarra, M. S., & Rodríguez, G. (2010). An approach to the dominant discourse of learning assessment in higher education. *Revista de Educación*, 351, 385–407.

Izci, K., & Caliskan, G. (2017). Development of Prospective Teachers' Conceptions of Assessment and Choices of Assessment Tasks. International *Journal of Research in Education and Science, 3*(2), 464-474.

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.

Jessop, T., & Tomas, C. (2017). The implications of programme assessment patterns for student learning. *Assessment & Evaluation in Higher Education*, *42*(6), 990-999.

Joy Cumming, J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in education: Principles, policy & practice*, *6*(2), 177-194.

Lipnevich, A. A., Guskey, T., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment and Evaluation: Principles, Policy, and Practice, 1-21*.

López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: a review of international literature. *Sport, Education and Society, 18*(1), 57-76.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational measurement: Issues and practice, 22*(4), 34-43.

Meyer, L. H., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., & Johnston, P. M. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly, 64*(3), 331-350.

Mohammadi, S., & Karupiah, P. (2020). Quality of work life and academic staff performance: a comparative study in public and private universities in Malaysia. *Studies in Higher Education*, *45*(6), 1093-1107.

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.

Panadero, E., Brown, G. T. L., & Courtney, M. G. R. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice, 21*(3), 365-383. doi:10.1080/0969594X.2014.919247

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education, 32*(1), 133-156. doi:10.1007/s10212-015-0282-5

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.

Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation, 38*(3-4), 84-92.

Quesada-Serra, V., Rodríguez-Gómez, G. & Ibarra-Sáiz, M.S. (2016). 'What Are We Missing? Spanish Teachers' Perceptions of Their Assessment Practices.'. *Innovations in Education and Teaching International 53* (1): 48–59.

Quinlan, K. M., & Pitt, E. (2021). Towards signature assessment and feedback practices: a taxonomy of discipline-specific elements of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 1-17.

Sambell, Kay, Liz McDowell, and Catherine Montgomery. (2012). *Assessment for learning in higher education*. London: Routledge.

Sandelowski, M. (2010). What's in a name? Qualitative description revisited. *Research in nursing & health*, *33*(1), 77-84.

Schuwirth, L. W., & Van Der Vleuten, C. P. (2019). Current assessment in medical education: Programmatic assessment. *Journal of Applied Testing Technology, 20*(S2), 2-10.

Silió, E. (2019). "La carrera de Matemáticas se dispara en plena era del Big Data". *El País*. https://elpais.com/sociedad/2019/05/28/actualidad/1559060134_280031.html

Stiggins, R. (2007). Assessment through the student's eyes. *Educational leadership*, *64*(8), 22.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of mixed methods research*, *1*(1), 77-100.

Tian-Ci Quek, T., Tam, W. S., X Tran, B., Zhang, M., Zhang, Z., Su-Hui Ho, C., & Chun-Man Ho, R. (2019). The global prevalence of anxiety among medical students: a meta-analysis. *International journal of environmental research and public health*, *16*(15), 2735.

Trenholm, S., Alcock, L. & Robinson, C. (2015). An Investigation of Assessment and Feedback Practices in Fully Asynchronous Online Undergraduate Mathematics Courses. *International Journal of Mathematical Education in Science & Technology 46*(8): 1197–1221. doi: 10.1080/0020739X.2015.1036946.

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic
assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher
Education*, *43*(5), 840-854.

Wächter, B. (2004). The Bologna Process: Developments and Prospects. *European Journal of
Education 39* (3), 265–273.

Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of
assessment: A phenomenographic study of the views of Hong Kong and Swedish
teachers. *Instructional Science, 33*(4), 283-309.

Wiliam, D. (2011). What is assessment for learning?. *Studies in educational evaluation*, *37*(1),
3-14.

Wiliam, D. (2019). Why formative assessment is always both domain-general and domain-
specific and what matters is the balance between the two. In H. Andrade, R. E. Bennett,
& G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 272–
294). Routledge.

Wu, Q., & Jessop, T. (2018). Formative assessment: missing in action in both research-intensive
and teaching focused universities?. *Assessment & Evaluation in Higher
Education*, *43*(7), 1019-1031.

Yan, Z., & Brown, G. T. (2021). Assessment for learning in the Hong Kong assessment reform:
A case of policy borrowing. *Studies in Educational Evaluation*, *68*, 100985.

CHAPTER 4

Assessment design decisions in practice: Profile identification in approaches to assessment design.

Javier Fernández Ruiz[1], Ernesto Panadero[2, 3], Daniel García Pérez[4] & Leire Pinedo[2]

Author Note

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

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

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

[4] Departamento de Psicología. Universidad Europea de Madrid

Authors ORCID:

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 to not having any conflict of interest regarding this manuscript. This research has been approved by the ethics committee from Universidad Autónoma de Madrid.

Abstract

This study aims to identify different profiles of higher education teachers based on the way they design their assessment methods. It also explores differential characteristics in each profile and differences in the assessment methods preferred by the teachers and those implemented in their subjects. Sixteen teachers from four universities participated carrying out a think-aloud simulation task: designing the assessment methods for a set of learning outcomes. Teachers' testimony during the task was transcribed and categorized using content analysis and an open-coding procedure. Three different patterns were identified: (a) focus on the feasibility of the assessment tasks, (b) on the alignment with the learning outcomes, or (c) alignment with teaching methods. Most of the participants focused only in one of the three elements. Teachers also designed different assessment methods in the simulation task in comparison with the ones they use in their subjects, despite the guidelines of assessing the same learning outcomes. A lack of resources is claimed as the reason of these differences. Implications for future research are discussed.

Introduction

Much has been studied on how different assessment practices can promote beneficial effects on students (Boud & Falchikov, 2006; Andrade & Heritage, 2013; McMillan & Moore, 2020), but educational practices are not usually consistent with empirical knowledge about assessment (Ibarra-Sáiz & Rodríguez-Gómez, 2010). In higher education, teachers can make decisions about several aspects of their assessment practices, and yet not much is known about the teacher's perspective when designing and implementing those practices (Bearman et al. 2017). The studies that address this issue have mostly focused on self-reported data (Bennett et al. 2011; Norton et al. 2013; Bearman et al. 2017), and literature exploring the assessment design while it takes place is missing. A better understanding of assessment design processes would be beneficial to increase the consistency between the wide empirical knowledge about assessment and the practices taking place in the classroom. Using an innovative data collection and analysis methodology, this study aimed to shed more light on the way in which university teachers design their assessment methods.

Theoretical framework

Assessment design in higher education

There are many studies discussing how specific assessment practices, if done properly, can have beneficial effects for students on aspects such as self-regulation (Panadero & Romero, 2014), motivation (Cauley & McMillan, 2010), and learning (Black & Wiliam, 1998). Although there is agreement on the importance of assessment as a promoter of learning, its implementation in higher education is sometimes far from what we know as formative assessment (Wu & Jessop, 2017). Studies such as Panadero et al. (2019) and Jessop & Tomas (2017) show a traditional assessment panorama in European universities. This is not necessarily due to a lack of effort on the part of

university teachers, who spend a third of their time designing and promoting assessment activities (Izci & Caliskan, 2017). However, it has been argued that, due to an inadequate understanding and use of assessment techniques, higher education teachers do not end up obtaining benefits in relation to teaching and learning (Stiggins, 2007).

To fully understand this phenomenon, it is necessary to investigate the way that teachers design their assessment practices, to identify challenging areas that may be stopping them from achieving a set of assessment practices they feel comfortable with (Bearman et al. 2017). The empirical understanding of this process would be able to lead to professional actions aiming to increase assessment effectiveness and make it more beneficial for teachers and students.

In this vein, authors such as Carless (2015) consider assessment design as one of the most critical aspects of assessment practices. It is known that this process can be extremely complex for teachers, balancing multiple tensions at the personal and institutional level. McDonald & Joughin (2009) and Meyer et al. (2010) have produced conceptual works representing the factors influencing assessment design. These factors go from general government and institutional policies to student and teacher characteristics. Several authors agree that assessment design by teachers is characterised by a tension between contextual and personal factors (McMillan, 2003; Bearman et al. 2017).

On one hand, contextual influences may be related to the subject, the discipline (Meyer et al. 2010), the learning outcomes, and the requirements of the department. The organizational culture is also frequently mentioned because teachers from the same faculty or department end up adapting their assessment practices to fit those of their peers (Carless, 2015; Bearman et al. 2017).

On the other hand, personal factors such as past experiences (Panadero, Brown & Courtney, 2014), backwash effects (Watkins et al. 2005), teaching experience (Quesada-Serra et al. 2016), and conceptions about assessment (Fernández Ruiz & Panadero, 2020) regulate assessment design. Brookhart (2011) recommends a series of skills for teachers in relation to the design of assessment practices, which include being able to articulate clear learning objectives, communicate to students what achievement of a learning objective looks like, understand the purposes and uses of the range of available assessment options and be skilled in using them, and having the skills to analyse classroom questions, test items and performance assessment tasks.

Research, however, shows a different picture in higher education assessment practices. The assessments implemented are often traditional and out of step with current trends (Jessop & Tomas, 2017; Panadero et al., 2019), and teachers do not seem to be able to justify the assessment practices they have designed (Postareff et al. 2012). Quesada-Serra et al. (2016) show how teachers, despite showing their agreement regarding the importance of formative assessment practices, do not use them in their classrooms with the same frequency (Norton et al. 2005).

These discrepancies are an indicator of teachers' difficulties when it comes to achieving coherence between their aims and their practices regarding assessment. It is therefore conceivable that, at some point in the design, teachers encounter certain challenges that they are not capable of facing, and that prevent them from carrying out assessment practices as they would like.

Assessment design processes

Although the factors affecting assessment design have been a relatively common research topic during recent years, much less is known about the design process itself.

Bearman et al. (2017) categorised assessment design actions into three different types: essential, selective, and meta-design.

The essential design activities are those actions which must be undertaken in order to design and implement an assessment task. These actions can be, for example, the development of an assessment task, feedback processes, and alignment with the teaching methods. Selective activities have to do with the grading processes, the control over contract cheating, or the fairness of the assessment practices. Finally, meta-design activities refer to the activities that educators undertook to manage their own design processes across iterations. Examples of this type would be the periodic review of assessment practices, collaboration with other teachers, or the active search for feedback from students.

In summary, it is known that educational assessment is a major key to ensure student achievement in higher education. However, the assessment practices used at this stage tend to be different from what the teachers would like to implement, which makes it necessary to investigate more carefully how the assessment design process is carried out. Assessment design is an extremely complex action, and the variety of different factors involved might make it difficult to recall in a further interview. Therefore, it is important to explore this process while it is taking place. This study aimed to investigate the way in which university teachers design their assessment practices using data collection methods new to this topic based on the observation and analysis of a simulated assessment design. To do so, we start from three research questions:

RQ1 – Are there different teacher assessment design profiles?

RQ2 – Are there differential characteristics in the teachers belonging to each of the profiles?

RQ3 – Which are the differences between the assessment practices designed during the assignment and the assessment practices implemented in their subjects?

Method

Participants

The participants were 17 higher education teachers, belonging to the branches of sport sciences, medicine, and mathematics. Two additional teachers were discarded from the analysis as they offered incomplete results. The data regarding academic discipline, years of experience, and gender is listed in the Table 4.1.

Table 4.1.

Information about the participants.

	Gender	University	Faculty	Teaching experience (years)	Current unit experience (years)
1	Male	1	Sport Sciences	31	16
2	Male	1	Sport Sciences	20	8
3	Male	1	Mathematics	20	3
4	Female	4	Sport Sciences	6	4
5	Male	2	Mathematics	31	1
6	Teaching pair (male + female)	1	Medicine	15	3
7	Male	1	Mathematics	28	2
8	Male	3	Sport Sciences	17	13
9	Female	2	Medicine	15	5
10	Male	1	Mathematics	28	1
11	Female	4	Sport Sciences	10	7
12	Male	2	Medicine	38	25
13	Female	1	Mathematics	6	2
14	Male	2	Mathematics	20	6
15	Male	4	Sport Sciences	14	11
16	Male	4	Sport Sciences	10	7

Procedure

The present study used a qualitative approach to explore the assessment process in a practice-based setting. It sought to describe and explain the phenomenon within a

professional context. To cover the range of data needed in this study and to obtain an intensive focus of the data, the sample size was limited.

Participants were contacted by email after analysing their syllabus as part of a previous study (see Fernández Ruiz et al., submitted for publication), which intended to explore and compare the normal assessment patterns in each faculty. After this analysis, the research team contacted the participants, having the focus on a balanced sample in terms of teaching experience and academic discipline. An individual session was held which included an initial interview and a simulation task. Before starting with the data collection, a pilot study was performed with three participants not included in the final sample. The goal was to ensure that the interview questions and the simulation task were understandable and to anticipate possible questions by the final participants.

During the initial interview, participants were asked about contextual factors such as teaching experience and training, and about the development of their assessment practices over the years. The questions asked were the following:

1. Why did you choose these assessment practices instead of any others?

2. Did you carry out modifications in the assessment practices of this subject? Which were the reasons for those modifications?

3. How did your approach to assessment changed with the pass of the years?

Once the interview was completed, the participants were informed about the characteristics of the simulation task. The aim of the task was to show as faithfully as possible how teachers design their assessment practices. Each teacher was given a set of contents and learning outcomes, taken from the syllabus of their actual subjects to ensure that they were familiar to them. Then, the researcher asked them to design an assessment methodology to assess 'in the best way possible' these learning outcomes.

Participants were informed that there were no correct or incorrect ways to perform the task, and that replicating the assessment methods they use in their subjects was a valid option. A thinking aloud protocol procedure was followed, and the participants were instructed to make explicit all the ideas and thoughts that they had during the process. The task was generally well-received among the participants, as the pilot study helped to anticipate their possible concerns about it (e.g., Do I have to use different assessment methods in this task? What if I would consider different approaches depending on the classroom size or characteristics?). The reflections of the teachers were audio recorded and transcribed for later analysis. The teachers took between 15 and 30 minutes to complete the task.

Data analysis

All the interviews were transcribed and coded using a content analysis approach. For RQ1, an open coding process was followed for the different actions that teachers performed during the task. After the analysis of the presence and order of the actions, three different profiles were found and described. Using the profiles as closed categories, two researchers independently read the transcriptions and coded the teachers into one of the profiles. This process had an initial agreement percentage of 90%, and subsequently the cases with discrepancies were discussed until they reached an agreement. Descriptive statistics were performed for RQ2, including frequencies, means, and standard deviations, and an open coding process was used for the interview responses. For RQ3, a closed coding process was followed using the categorization by Bearman et al. (2017). Software tools Atlas.ti 8 and SPSS 26 were used during the data analysis stage.

Results

RQ1 – Are there different teacher assessment design profiles?

Actions performed by the teachers during the task were coded into eight categories, described in Table 2.

Table 2.

Assessment design process categorization.

Evidence	Teachers selected one assessment evidence.
Grading	Teachers reflected on how the evidence must be graded, its weight on the students' final grade, and the minimum score required for passing.
Learning outcomes	Teachers read and analysed the learning outcomes provided and explored possible links with their assessment practices.
Teaching methods	Teachers selected one or several teaching methods and explored possible links with their assessment practices.
Feedback	Teachers reflected on how feedback must be provided to students.
Rubrics	Teachers explore the use of rubrics as a tool for students or themselves.
Contract cheating	Teachers explore the possible ways of preventing contract cheating in their assessment practices.
Students' participation	Teachers explore the possible ways of implementing students' participation in their assessment practices, mostly via self- or peer assessment.

Every action may have been carried out by the teacher once or more times during the task, and each time it would be coded independently. The presence or absence of the actions and the order in which they are carried out were used to find possible profiles among the participants. After the content analysis of the teachers' actions, three different profiles were seen in the design of the assessment practices. We called these profiles Classic, Competence, and Cohesive.

Classic profile

A total of 7 teachers were categorised in the classic profile (Figure 4.1), designing a total of 25 instruments. In this profile, teachers designed their assessment practices in the most minimalist way possible. The teachers of this profile prioritised the efficiency and logistics of the assessment practices above all else. During the task, they tended to think directly about the assessment instruments they considered most convenient, without

placing too much emphasis on the competences assessed by each one of them. For each instrument, they usually indicated aspects such as the grading processes (common in all profiles), the temporality, the security regarding contract cheating, and the available resources.

Figure 4.1. Classic profile flowchart.

Note: Numbers inside the boxes represent how many occasions the category was used. Numbers in the arrows represent how many occasions the design process went from one category to the next one.

As shown in Figure 4.1, 25 assessment instruments were designed in this way. Most of the teachers went on to give them a percentage of the grade, without further consideration. On two occasions they questioned the best way of preventing contract cheating, and on two other occasions whether they could involve the students in some way.

> *It depends on how you want to name it, but it would be reasonable that the partial examinations have a grade. When that grade exceeds the threshold, we are going to make, then the subject is passed. What would I do? According to my experience*

these years… to make the weight of each partial examination progressive. The

first one will be 25% and so on until we have 100%.

Teacher 3 - Mathematics

Competence profile

A total of 5 teachers were categorised as competence profile, designing a total of

15 assessment instruments as shown in Figure 4.2. The teachers of this profile seemed to

give priority to the coherence between the contents and learning outcomes to be assessed,

and the chosen assessment instruments.

Figure 4.2. Competence profile flowchart.

Note: Numbers inside the boxes represent how many occasions the category was used. Numbers in the arrows represent how many occasions the design process went from one category to the next one.

The most common pattern of this type of teacher was to take a close look at the

learning outcomes of the subject, reflect on the best way to elicit them in an assessment

task (either the one they currently use or some other proposal), and describe the

assessment task they would use and how they would grade it. However, it is relatively

common for them to make a pause to consider the feedback provided or the use of rubrics.

If we want the students to be able to collect and interpret relevant data to make

judgements [one learning outcome from her subject] … the assessment tasks

would be the ones I am using right now, an assignment and a final examination.

That makes them learn how to search information, or even using the information

they already have in their notes, and make a judgement, a reflection, to answer

one particular question.

Teacher 11 - Sport Sciences

Cohesive profile

A total of 4 teachers were categorised as cohesive, designing a total of 10 assessment instruments. Teachers of this profile seemed to pay special attention to integrating their assessment practices in a broader context. Normally, they tended to think first of the teaching methods that they would implement in their subject, after which the chosen assessment practices were a logical consequence. Teachers in this category are also characterised by being especially attentive to coordination between subjects, or even between different academic courses.
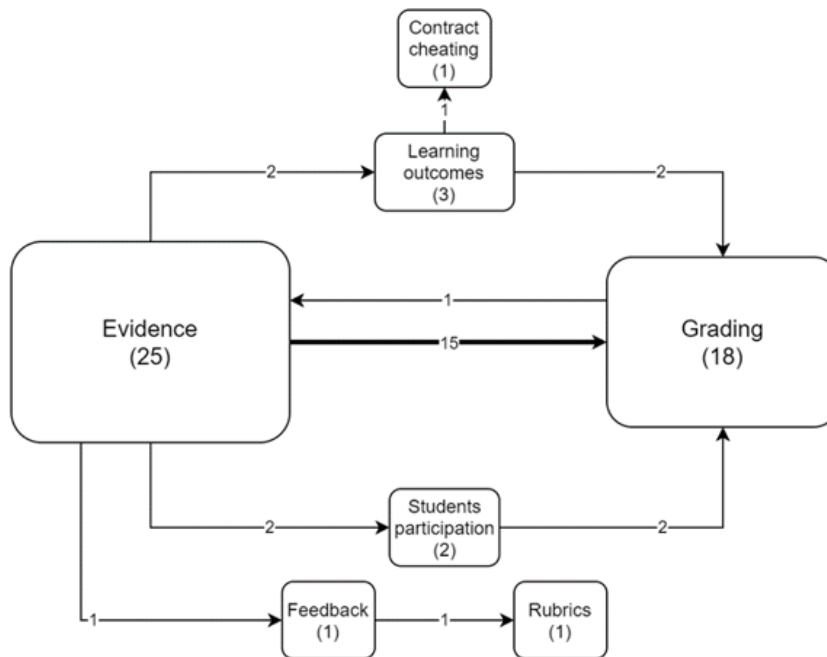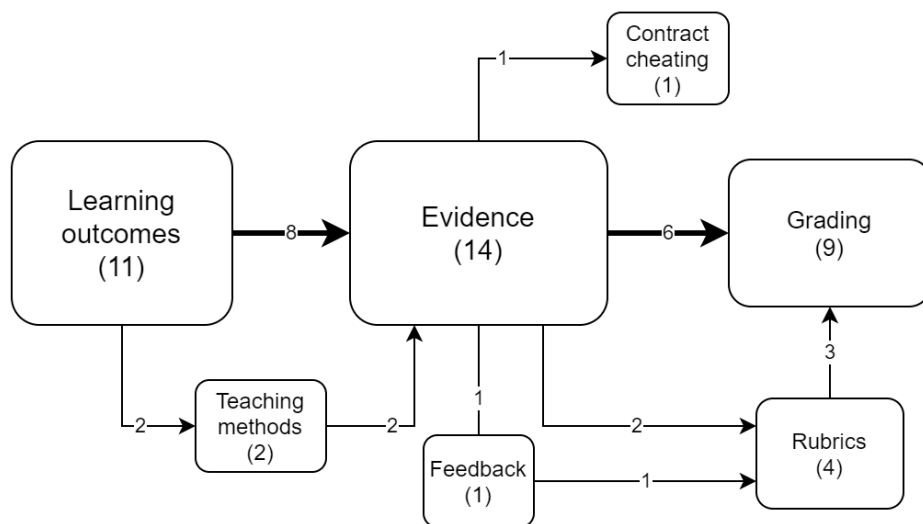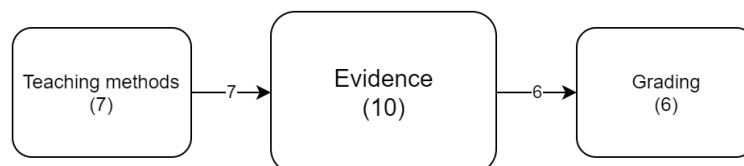

Figure 4.3. Cohesive profile flowchart.

Note: Numbers inside the boxes represent how many occasions the category was used. Numbers in the arrows represent how many occasions the design process went from one category to the next one.

My teaching methodology is based on case-solving and problem-based learning

(..). I would link to that one skills seminar, where the student can show their skills.

It does not have to be only clinical exploration, but calculus, critical reading, and

writing as well. All of it would be incorporated into the assessment element.

Teacher 12 - Medicine

The syllabi of each teacher's subjects were analysed to explore possible differences in the assessment practices carried out by teachers in each profile. There were differences between the teachers of the competence profile and the rest, as shown in Table 4.3. Their assessment practices showed a continuous and practical approach, highlighted by the use of practical exams and a greater use of classroom practices. In addition, it was common for teachers of this profile to evaluate attendance systematically.

Table 4.3.

Assessment practices in each profile.

	N	FE	PrE	PE	PO	AS	PR	AT
Classic	7	5	0	3	2	3	3	2
Competence	5	3	3	3	0	3	5	4
Cohesive	4	3	0	0	0	0	3	1

Note: FE = Final Examination, PrE=Practical examination, PE=Partial Examination, PO=Portfolio, AS=Assignments, PR=Practices, AT=Attendance

RQ2 - Are there differential characteristics in the teachers belonging to each of the profiles?

Assessment training

This research question searched for differences in the academic domain, teaching experience, in-service training, and assessment conceptions and practices among the teachers in each profile. In relation to the academic domain, there were no appreciable differences, with representatives of the three profiles in each of the degrees. There were differences, however, in the teaching experience of each profile, as shown in Table 4.4. Specifically, the teachers categorised in the competence profile seemed to be those who had less experience, for both in teaching in higher education and teaching in their current subjects.

Table 4.4.

Degree and experience of teachers in each profile.

	N	Teaching experience (years)	Current subject experience (years)
Classic	7	22.3 (SD = 7.52)	7.1 (SD = 6.12)
Competence	5	13.8 (SD = 8.56)	4.6 (SD = 2.30)
Cohesive	4	21 (SD = 13.11)	10.3 (SD = 10.14)

Regarding in-service training, competence teachers were also the ones who did more teacher training courses, despite being the less experienced teachers. As shown in Table 4.5, teachers from this profile had done a total of 14 courses, mostly related to teaching and assessment practices.

Table 4.5.

Teaching training courses by teachers in each profile.

Topic	Classic	Competence	Cohesive
Assessment methods and tools	3	3	3
Teaching methods	3	5	0
Innovation	1	2	2
Use of ITCs	1	1	1
Syllabus elaboration	0	2	0
Interaction with students	0	1	0
Average per teacher	**1,14**	**2,8**	**1,5**

Assessment conceptions and practices

Teachers in the three profiles showed different trends in the way they have developed their assessment practices over time. The first question they were asked in this regard is why they had opted for the methods they currently use in their subjects. The teachers of the classic profile did not offer a clear explanation, and their answers indicated a certain inertia with respect to what was the tradition in their department or their faculty.

A no-exam subject… it would be a bit weird. The student congress [an assessment instrument consisting in students' presentations simulating a scientific conference] is basically inherited from an old subject of the previous plan.

Teacher 9 - Medicine

Over the years, they have tried to give greater weight to continuous assessment, but without modifying the set of instruments they had established. Instead, they varied the weight of each instrument on the final grade to make students focus more on those related to continuous assessment.

The conceptions that these teachers had about assessment had also changed over time. They claimed to rely on the students' responses to check whether their assessment practices were working or not. These answers had led them to reflect on the meaning of the final exam, and to consider alternative assessment practices. However, they were reluctant to make meaningful modifications of their assessment practices, and most of them argued departmental requirements as a constraint.

> *Regarding the assessment criteria, each teacher in his group has freedom. But that freedom exists up to a certain point (…) There are aspects in which the teacher cannot… I do not think he could say… "well, I'm not going to make exams".*
>
> *Teacher 5 - Mathematics*

The teachers of the competence profile generally declared that their assessment practices were designed with the aim of preparing their students for their professional future, trying to give them a practical and competence approach. However, they stated that they were limited by the time that a formative and continuous assessment requires, which they considered sometimes overwhelming. Regarding their development over the years, these teachers had also focused on promoting continuous assessment, but unlike the classic profile, with much more freedom to modify their assessment practices. In relation to their conceptions of assessment, they showed a growing interest in preparing their students for their professional future, adopting a practical and competence approach.

In addition, they were aware of the changes that student cohorts have undergone since they began teaching, and they tried to adapt to them.

Our students' brains now work in a different way from when I was studying. Because society is changing, and things must be immediate, they have to be a click away. And then I think that sometimes we change things so that they have it just one click away, so they actually read it. And of course, if it costs them more than that, we already limp.

Teacher 4 - Sport Sciences

The teachers of the cohesive profile declared that the assessment of their subjects had been a logical result of their teaching methods, or the characteristics of the subject. These teachers did not offer a concrete answer about the development of their assessment practices over the years.

I have a colleague with whom I have a good relationship. So, I knew that he did project-based learning. I have to tell you; at that time, I did not have much faith in it. Well, we talked, we had coffee and I asked him how he did it. So, well, I started going to his classes ... in fact I think that that course I went to many classes to see how he did it ... And well, I saw the dynamics of that project-based learning, so that was one of the influences to teach using project-based learning.

Teacher 2 - Sport Sciences

RQ3 – What are the differences between the assessment practices designed during the assignment and the assessment practices implemented in their subjects?

Given that the set of contents and learning outcomes to be assessed during the task was the same that the teachers must assess in their subjects, similarities and differences between the assessment practices designed during the task and those of their subjects were analysed. However, 10 teachers (62.5%) implemented changes in their assessment

practices during the task. According to the categorization of Bearman et al. (2017), seven teachers carried out "Essential design activities" and three teachers carried out "Selective design activities". Regarding the first group, the changes proposed consisted of creating additional tasks (n = 4), eliminating existing tasks (n = 1), or changing their format (n = 2). The second group modified the percentages of the assessment tasks already implemented in their subject.

When asked about the reasons between the differences in the practices designed in the task and those of their subjects, most of their answers indicated that implementing these methods in real practice would be excessively difficult for several reasons. Among them would be the workload it would entail for them (n = 3) or for their students (n = 1), logistical or coordination difficulty (n = 2), and the ratio of students to teachers (n = 4). Interestingly, two teachers stated that the only reason they had not implemented these methods in their subject is because they had not thought about it.

> *I think the time to do it has not yet come for me. But I do believe that it is being generated little by little ... because the assessment must be in tune with the teacher, and no matter how much they want to innovate, or want to do, if they do not feel safe doing it, or have their own ... fears, and sometimes training, which we also sometimes lack regarding assessment.*
>
> *Teacher 15 - Sport Sciences*

To summarise, this study presents three different profiles regarding assessment design, which have been called classic, competence, and cohesive. Each profile has differential characteristics in their approach to assessment design.

Discussion

The aim of this study was to explore the way in which higher education teachers design their assessment practices. The results provide new information as they are based on simulated performance instead of self-reported data.

The first research question explored whether it was possible to identify different profiles in the design of assessment practices. The results show different approaches regarding assessment design, which have been called classic, competence, and cohesive profiles. It should be noted that there seems to be a certain consistency in the design of assessment tasks, as the participants designed all their assessment tasks following the same pattern.

The classic profile was the most common among the participants. These teachers do not make considerations prior to selecting the instrument to be used. Choosing a determined set of instruments is the first step in their assessment design process, ending later with stating the grading weight of each one. Few teachers in this profile make additional considerations.

It is remarkable that many of the teachers in this profile showed a more formative approach to assessment in the initial interview questions than in their simulation task responses. It may imply that, even if they acknowledge the value of formative assessment, they lack the abilities to translate it to their assessment practices. This issue has been addressed in several studies (Norton et al. 2010; Smith, 2011), and a variety of reasons might explain it. Participants of this study argue that they are influenced by the traditional assessment practices in their faculty or department, a well-known issue regarding assessment design (MacDonald & Joughin, 2009). The lack of training and resources to change their assessment practices is also another possible explanation.

The predominance of this profile might be in line with results such as those of Ibarra-Sáiz & Rodríguez-Gómez (2010) or Panadero et al. (2019), which show a

traditional approach to university assessment. If most of the teachers do not reflect on formative aspects during their assessment design, or lack the resources to put them in practice, it is to be expected that they perpetuate summative approaches, as has been the tradition until now.

The competence profile was next in use. The teachers in this profile design their assessment practices starting from the learning outcomes of the subject, reviewing them in detail and reflecting on what would be the best way to ensure that students have achieved them, while helping them to do so. These teachers also tend to reflect on the feedback offered and more innovative methodologies are proposed.

It is noteworthy that the teachers of this profile were also the least experienced, both in university teaching in general, and in their current subjects. The trend towards a more formative approach among the youngest teachers has been argued by studies such as by Quesada-Serra et al. (2016) and seems to be supported by our results. However, teachers in this profile were also the most trained ones, in terms of in-service training courses. It is argued that the importance of pre-service training programmes is to develop assessment skills among teachers (Picos & López Pastor, 2013). Our results also show the need of designing in-service training courses oriented to providing higher education teachers with the ability to design a formative assessment methodology.

Teachers in the competence profile, however, also faced challenges regarding their assessment design. They declared having a hard time balancing their assessment practices, which are often very complex, with their available resources, in terms of workload and external help. Much has been studied regarding the efficacy of formative assessment in terms of student learning, but these results raise questions about its efficiency. Authors such as Higgins, Grant & Thompson (2010) detail several principles for increasing formative assessment efficiency, including strategic curriculum review, use

of IT, self- and peer assessment, oral feedback, in-class assessment, and group assessment. Apart from the latter two, none of them were mentioned by the participants of this study.

Finally, the cohesive profile is based on coherence between teaching and assessment practices. Teachers in this profile were not limited to designing independent assessment practices, even if the simulation task pointed in that direction; rather, their design went further by including various aspects of the teaching-learning process. They usually started the process by raising the teaching methods, starting for example from project- or problem-based learning methodologies. Once this was done, their assessment practices were a natural consequence of the teaching methods employed.

Teachers in this profile showed a less structured thinking process regarding assessment practices themselves, as they spent most of the time outlining the subject in general. They were also the ones who had been teaching their current subjects the longest, which may imply that some experience and trial-and-error processes are necessary to get to this point. Previous research exploring thinking processes in novice and experienced teachers seems to support this idea. Hall & Smith (2006) argued that experienced teachers use less structured and more implicit thinking processes when planning their subjects. It also seems to be the case in the present study.

The last research question explores the changes between the assessment practices designed during the task and the assessment practices implemented in the actual teachers' subjects. Since the task was a mere simulation, with no necessary consequences in practice, the comparison between the assessment practices designed during it and those implemented in their subjects is appropriate to explore further the limitations of the assessment intended by the teachers when it is carried out in actual practice.

Implications for professional practice

It is noteworthy that most of the teachers designed different assessment practices compared with the ones they use in their subjects, even though the aim was to assess the same set of learning outcomes. Most of the changes consisted of broadening the assessment practices, adding new tasks, or modifying existing ones to make them more complex. The main reason why they decided not to use these methods in their classrooms refers to the lack of resources, or the workload that it would imply, both for their students and themselves. This discovery aligns with previous studies such as those by Postareff et al. (2012), Norton et al. (2013), or Quesada-Serra et al. (2016), where teacher testimony showed many more formative approaches than there were put in practice.

The discrepancies between intentions and practice go back a long way. Norton et al. (2010) and Smith (2011) already showed that, even if teachers are aware of the formative purposes of education and agree with them, they are usually not capable of putting them into practice. Most of the participants recognise that their assessment practices are not ideal, but rather an approximation that is achievable through the available resources.

The definition of 'resources' also varied among the participants of this study depending on their profile. Classic-profile teachers seemed to lack the knowledge and ability to make meaningful innovations to their assessment practices in practice, and they tended to use assessment practices coherent with the tradition in their department. Competence-profile teachers might have had the knowledge, as they were usually well-trained regarding teaching and assessment. However, even if they considered themselves capable of designing an effective set of assessment practices, they claimed to be unable to balance them and make them efficient in terms of workload and coordination.

Both challenges mentioned by the participants are relatively well-known among the assessment literature. Based on previous theory and research work, we offer a series

of recommendations for institutions and teachers that aims to help them face the challenges regarding assessment design.

Teacher training and organizational solutions

In the country where our study was based (Spain), pre-service pedagogical training is not compulsory for higher education teachers. This results in many teachers starting their assessment design processes without any literacy or theoretical knowledge about how assessment must be done. This makes it especially important to offer high-quality in-service training, to provide them with the empirical and practical tools to improve their assessment practices.

The previous literature provides some instructions for in-service training about assessment. First, authors such as Brookhart (2011), based in the seminal 'Standards' framework (American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990) offers a comprehensive list of assessment-related competences that teachers must acquire. Some examples of these competences are understanding learning in the content area they teach, having a wide repertoire of assessment strategies, or providing effective and useful feedback. For more information, see Brookhart (2011. p.7).

Nevertheless, a set of training courses covering the competences proposed by Brookhart can still be unsuccessful if not implemented correctly. Xu & Brown (2016) carried out a systematic review on assessment literacy, concluding that teacher education must address four pertinent issues to be effective for teachers.

First, teachers must have a solid assessment literacy knowledge base, implying that teacher education programs must include assessment courses as part of their curriculum (Sato et al., 2008). Second, assessment education needs to be long enough to let teachers acquire a deep understanding about assessment. It also needs to establish

connections between assessment theory and practice (Lyon, 2013). Third, it needs to address teachers as individuals and professionals (Hill et al., 2010). Every teacher has a background of conceptions, expectations, and prior experiences about assessment. Lastly, teacher educators need to understand that assessment literacy development is not merely an accumulation of assessment knowledge, but rather the development of a sophisticated, contextually appropriate set of inter-related competencies (Xu & Brown, 2016).

It is also important to consider that in-service teachers have limited time and opportunities for pedagogical training, so institutions must adapt to these circumstances. On-line learning (Fan, Wang & Wang, 2011) and reflecting on their assessment practices (Smith, 2011) can be methods for them to learn without compromising their workload.

Time management, individual solutions

The other big issue mentioned by our participants was the lack of time to put in practice their preferred assessment methods. Efficiency in formative assessment has been an underexplored topic in the literature, but authors such as Higgins et al. (2010, p. 10) provide useful strategies for carrying out high-quality assessment methods without compromising the time available to the teacher. These strategies are as follows:

1. Strategic curriculum review: The assessment workload can be reduced by avoiding repetition. Continuous assessment is a requirement in many European universities, but there are methods to minimize the total number of products assessed. These methods could include exempting students from a final assessment based on coursework performance (Hornby, 2005, p.22).

2. Use of ICT: Using automatically assessed virtual tools is a way to obtain general knowledge of your students' progress without having to read and assess individual tasks.

3. Group assessment: It can mean less workload for the teacher while different skills are developed by the students (Higgins et al. 2010).

4. In-class assessment: can include possible periodic tests, assessing each other, and gaining exemption from the final exam (Hornby, 2005, p.23). Additionally, oral informal feedback given during in-class activities can be easily combined with the use of ICT, and it is especially useful in large classes.

5. Peer and self-assessment: its impact on learning and self-regulation is one of the main lines of work on educational research these days (Panadero et al., 2016). It is also easily combined with the use of all the previous categories. Apart from a method to reduce teacher workload, its use can be enormously beneficial for students, if implemented correctly.

Implications for future research

These results provide a detailed view of what is happening when assessment is designed. Until now, research had focused on the factors, both personal and contextual, that influenced the process (McMillan, 2003; Bearman et al. 2017). However, despite the broad conceptual work published on this topic, there were few empirical studies exploring how this process is carried out in practice. The discovery of three different profiles is a great leap in research on assessment design in higher education. These results open a new avenue of research, with relevant questions for research that must be answered in subsequent studies.

The first question is if it is possible to identify predictors of the different profiles. One aim of the present study was to identify such predictors, and in our results, it was shown how teaching experience and in-service training may influence how teachers design their assessment practices. However, our small sample made it difficult to make solid claims regarding predictors.

Other factors such as disciplinary approaches, departmental culture, and student characteristics have been found to influence teachers' assessment design (Meyer et al. 2010; Bearman et al. 2017; Fernández Ruiz et al., under review). These variables could theoretically work as a predictor for the profiles found in this study. To find these predictors and to connect them with the profiles described here would be fundamental. This knowledge would allow teacher training programmes adapted to the personal necessities and approaches of each teacher and based on their real practices regarding assessment design, which is still a challenge for educators (Korthagen, Loughran & Russell, 2006).

The second question is about whether these approaches to assessment design remain stable throughout the teacher's professional career. Our results show a difference in the teaching experience among teachers in each profile. Less experienced teachers tend to show a formative approach in assessment design, and more experienced teachers are more inclined towards the competence and cohesive profiles. It is important to understand the cause of these differences.

It may be due to the new professional context, which would imply a heavier cognitive load as much of the processes have not been automated yet. In this case, once they have connected specific learning outcomes to specific teaching and assessment practices, they would focus on different factors during further redesigns, as was observed in the other profiles. This might be counterproductive, considering that both the learning outcomes and the social and professional environment can be enormously variable throughout a teacher's professional career. Teachers would then have difficulties to adapt their instructional and assessment practices to the current context.

On the contrary, the differences among teachers regarding experience can be an effect of a stronger scientific and institutional focus on formative assessment in recent

years. In this case, these differences would be considered a positive outcome of this new approach and it would be expected that teachers maintain the same approach throughout their professional careers.

The third question is about the possibility of combining several profiles. The profiles identified in this study show three different approaches to assessment design. Teachers seem to focus on available resources, learning outcomes, or teaching method alignment, and the combination of these is scarce. However, the three approaches are fundamental to achieve high-quality assessment practices. It is desirable that university teachers are able to design an assessment methodology aligned with their teaching methods and the learning outcomes, but also feasible and efficient considering their available resources. It is necessary to further explore which would be the best way to help the teachers to balance their assessment designs paying attention to the three factors.

Limitations

As the present study had a comprehensive and in-depth approach to data, the number of participants was smaller than survey type studies. Having a sample consisting of only 16 teachers may affect the significance of the findings and its applicability in other contexts. A significant effort was made to maintain a diverse sample in terms of institution (four universities), discipline, and teaching experience. Also, teachers from our study reported assessment design processes highly related with the constraints they face in their practice. Exploring the applicability of our results to teachers from different professional backgrounds must be done in further studies.

Conclusion

The present study shows an innovative research approach to investigate how university teachers design their assessment practices. We found three different design patterns: classic, which focuses on the feasibility of the assessment tasks; competence,

which focuses on alignment with the learning outcomes; and cohesive, which focuses on alignment with their teaching methods. It is striking that most of our participants carried out a design focused only in one of the three elements. Interestingly, the assessment practices that the participants designed in the simulation task were different from the ones they use in their real lectures. When asked about this point, participants explained these differences because they lack the resources needed to implement these complex practices. This study represents a starting point for practice-based research involving assessment in higher education. Our results make explicit the need to answer further questions to fully understand how higher education teachers can be helped to link their assessment practices with the vast empirical knowledge about assessment.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA). (1990). The standards for teacher competence in the educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30–32.

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

Bennett, S., Thomas, L., Agostinho, S., Lockyer, L., Jones, J., & Harper, B. (2011). Understanding the design context for Australian university teachers: Implications for the future of learning design. *Learning, Media and Technology, 36*(2), 151–167. doi:10.1080/17439884.2011.553622.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & evaluation in higher education, 31*(4), 399-413.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, *30*(1), 3-12.

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963-976.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The clearing house: A journal of educational strategies, issues and ideas*, *83*(1), 1-6.

Fan, Y. C., Wang, T. H., & Wang, K. H. (2011). A web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education*, *57*(2), 1727-1740.

Fernández Ruiz, J., & Panadero, E. (2020). Comparison between conceptions and assessment practices among secondary education teachers: more differences than similarities. *Infancia y Aprendizaje: Journal for the Study of Education and Development*, *43*(2), 309-346.

Fernández Ruiz, J., Panadero, E. & García-Pérez, D. (2021). Assessment from a disciplinary approach: Design and implementation in three undergraduate programmes. Manuscript submitted for publication.

Hall, T. J., & Smith, M. A. (2006). Teacher planning, instruction and reflection: what we know about teacher cognitive processes. *Quest, 58*(4), 424-442.

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.

Hill, M., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing assessment-capable teachers: What should preservice teachers know and be able to do?. *Assessment Matters*, *2*, 43-64.

Hornby, W. (2005). Dogs, stars, Rolls Royces and old double-decker buses: Efficiency and effectiveness in assessment. In: *Enhancing practice: Reflections on assessment: Volume 1,* 15–28. Gloucester: Quality Assurance Agency for Higher Education.

Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2010). An approach to the dominant discourse of learning assessment in higher education. *Revista de Educación*, *351*, 385-407.

Izci, K., & Caliskan, G. (2017). Development of Prospective Teachers' Conceptions of Assessment and Choices of Assessment Tasks. International *Journal of Research in Education and Science, 3*(2), 464-474.

Jessop, T., & Tomas, C. (2017). The implications of programme assessment patterns for student learning. *Assessment & Evaluation in Higher Education*, *42*(6), 990-999.

Korthagen, F., Loughran, J., & Russell, T. (2006). Developing fundamental principles for teacher education programs and practices. *Teaching and teacher education*, *22*(8), 1020-1041.

Lyon, E. G. (2013). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education*, *35*(7), 1208-1229.

Macdonald, R., & Joughin, G. (2009). Changing assessment in higher education: A model in support of institution-wide improvement. In G. Joughin

(Ed.), *Assessment, learning and judgement in higher education* (pp. 1–21). Rotterdam, Netherlands: Springer.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational measurement: Issues and practice*, *22*(4), 34-43.

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.

Meyer, L. H., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., et al. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly, 64*(3), 331–350. doi:10.1111/j.1468-2273.2010.00459.x.

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.

Norton, L., Norton, B., & Shannon, L. (2013). Revitalising assessment design: what is holding new lecturers back?. *Higher Education*, *66*(2), 233-251.

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.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133-148.

Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 365-383.

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*(4), 803-830.

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.

Picos, A. P., & López-Pastor, V. M. (2013). Haz lo que yo digo pero no lo que yo hago: sistemas de evaluación del alumnado en la formación inicial del profesorado. Do What I Say, Not What I Do: Student Assessment Systems. *Revista de educación*, *361*, 279-305.

Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, *38*(3-4), 84-92.

Quesada-Serra, V., Rodríguez-Gómez, G., & Ibarra-Sáiz, M. S. (2016). What are we missing? Spanish lecturers' perceptions of their assessment practices. *Innovations in Education and Teaching International*, *53*(1), 48-59.

Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, *45*(3), 669-700.

Smith, J. (2011). Beyond evaluative studies: perceptions of teaching qualifications from probationary lecturers in the UK. *International Journal for Academic Development, 16*(1), 71–81.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational leadership*, *64*(8), 22.

Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science*, *33*(4), 283-309.

Wu, Q., & Jessop, T. (2018). Formative assessment: missing in action in both research-intensive and teaching focused universities?. *Assessment & Evaluation in Higher Education*, *43*(7), 1019-1031.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

CHAPTER 5

A nationwide analysis of in-service training courses about assessment for higher education teachers.

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 to not having any conflict of interest regarding this manuscript.

This research has been approved by the ethics committee from Universidad Autónoma de Madrid.

Abstract

Starting from the theoretical models about assessment literacy, this study presents a nationwide analysis of in-service teacher training courses about assessment. Every teacher training course from Spanish public universities (N = 1627) was screened. Data about all available courses related to assessment (N = 82) was collected and analysed. Different approaches in terms of format and duration were found depending on the university. While some universities use a massive webinar approach to teacher training, others prefer longer and more intensive courses. Regarding courses' contents, 25 themes were found, grouped into 6 main thematic areas. Courses format and contents were compared with literature foundations on assessment training. Lastly, three quality criteria are proposed based on the theoretical models about assessment literacy, finding that only 3 of the 25 universities analysed would be offering a high-quality assessment training to their teachers.

Keywords: assessment literacy; teacher education; assessment training

A nationwide analysis of in-service training courses in assessment for higher education teachers

University assessment practices are diverse, with varying profiles depending on factors such as the academic discipline or the year level of the students (Lipnevich et al., 2021; Panadero et al., 2019). These assessment practices have direct effects on students' academic achievement, and it is therefore crucial to understand why and how the practices are chosen and implemented. In many countries, teachers are responsible for taking the decision to design and implement a particular set of assessment practices, so it is important to understand the factors that influence such a decision (Adachi et al., 2018; Bearman et al., 2017). One key factor to address is the type of training that higher education teachers receive in the area of educational assessment or, in other words, the training they receive in aspects such as how to deliver feedback, how to design exams, and so on.

Assessment literacy is considered to be one of the requirements that enables a teacher to carry out a high-quality assessment in higher education. According to Popham, "teachers, and there are no exceptions, need to understand the nature of the instruments being employed to judge them as professionals" (Popham, 2011, p. 269). The growing interest in assessment literacy is due to reasons such as the role of assessment in student learning (Black & William, 1998) and the importance of teachers as agents in educational assessment (Leung, 2013).

Importantly, pre-service training for university lecturers is scarce, or sometimes even non-existent (Beziat & Coleman, 2015; Popham, 2009). For this reason, offering teachers adequate in-service training is essential if they are to be able to carry out high-quality assessments. However, we have at the moment very limited evidence about how this in-service training is being delivered. This study will carry out an exploration at a

nationwide level to give a complete vision of how in-service training in assessment in higher education is implemented.

Assessment literacy models

Despite the importance of assessment literacy as a theoretical construct, studies show that university teachers struggle when translating their instructional goals into actual assessment practice (Fernández Ruiz et al., 2021; Norton et al., 2010; Smith, 2011). As many authors suggest, if teachers receive training in assessment then these struggles are significantly reduced or even do not occur, and there can be a remarkable improvement in the quality of assessment practice (DeLuca & Bellara, 2013; McMillan, 2017).

In this regard, it is necessary to understand what makes an assessment literate teacher. Pastore and Andrade (2019), in their review, summarized several definitions of assessment literacy that had been given over the years, and concluded that assessment literacy has become a very complex concept. From the most basic definitions that are mostly focused on practical actions (Schafer & Lissitz, 1987), assessment literacy is now understood as "teacher capabilities to plan and implement quality assessment tasks, to interpret evidence and outcomes appropriate to the assessment purpose and type, and to engage students themselves as active participants in assessment of their own learning" (Looney et al., 2017, p. 2). Authors such as DeLuca and Bellara (2013) have argued that assessment literate teachers understand how to construct, administer, and score reliable assessments and to communicate valid interpretations about student learning. Some authors during the last decade have developed this field, to the point of proposing theoretical assessment literacy models. This study aims to analyse the quality of in-service assessment training in Spanish public universities according to one theoretical

model of assessment literacy. As we will use it as a framework for our findings, we will expand on the TALiP model put forward by Xu and Brown (2016).

Teacher Assessment Literacy in Practice (TALiP)

In 2016 Xu and Brown presented their own assessment literacy model, which was strongly influenced by the work of DeLuca (2012) and Willis et al. (2013). Their framework consists of six components. The bottom of the pyramid covers the knowledge base, which is essential for effective assessment practice (Maclellan, 2004). The knowledge necessary for teachers to become assessment literate has been discussed in several previous works including the foundational "Standards" (AFT, NCME & NEA, 1990) and their later revision by Brookhart (2011). In TALiP, Xu and Brown identified seven main areas in which teachers must acquire knowledge.

At the second level, the knowledge is filtered and interpreted by the teachers' conceptions of assessment (Barnes et al., 2015; Fives & Buehl, 2012). Several studies have investigated the strong relationships between the conceptions and the practices of assessment (Brown et al., 2009; Fernández Ruiz & Panadero, 2020), and have shown that assessment practices are strongly influenced by teachers' conceptions. Importantly, cognitive and affective dimensions, views of learning and epistemological beliefs are also included in the model.

Despite the importance of these conceptions, in-service "teachers cannot do whatever they please in actual practice since they are employed within an immediate workplace community and larger social, political, and cultural contexts" (Xu & Brown, 2016, p. 157). These contextual variables have an effect on assessment practices, through policies, norms, rules, regulations, and conventions, to "create a culture of certainty and compliance that is not easily challenged by teachers" (Scarino, 2013, p. 312).

Accordingly, the next level of the model is oriented towards teachers' decision making. This is the process by which teachers balance the demands and constraints of those external factors against their own beliefs and values (McMillan, 2003). Teachers need to make compromises, as argued by Carless (2011), in order to perform appropriate assessment practices, as their conceptions and expectations about assessment are probably not fully suited to the context. Therefore, Xu and Brown (2016) argue that assessment literacy is better understood as teacher assessment literacy in practice (TALiP), which includes the various compromises that teachers make to reconcile these tensions.

The model also emphasizes the importance of teacher learning for advancing TALiP. Teacher learning is defined as "both a process of active individual construction and a process of enculturation into the … practices of wider society" (Cobb, 1994, p. 13). Accordingly, it is proposed that teacher learning can be achieved in two main ways: through reflective practice, and though participation in community activities.

Lastly, the TALiP model pursues the (re)construction of the teacher's identity as an assessor, as opposed to the traditional "instructor" identity. This change represents the ultimate goal of the model, as it can play an important role in provoking teachers' thinking, promoting dialogue and shared meaning, and shaping teachers' professional judgements (Xu & Brown, 2016).

The importance of assessment literacy is clearly acknowledged in the theoretical literature, but teacher training in higher education is not necessarily based on scientific foundations (Brown, 2015; Ibarra & Rodriguez, 2010; Palacios & López-Pastor, 2013). In-service training programmes in higher education demonstrate a series of pros and cons in how they translate assessment literacy models into real practices. In the next section these will be discussed.

In-service training programmes in assessment in higher education

When considering the training of teachers according to the models mentioned previously, it is important to differentiate between pre-service and in-service training. In-service training has specific characteristics that need to be considered when training programmes are designed. Two crucial characteristics are emphasized here. First, it is necessary to act in a limited time, since higher education teachers must already balance a significant pedagogical, research and management workload (Anderson, 2006; Langford, 2010). However, second, in-service training has an important advantage: it makes it possible for teachers to experience in real time the new concepts and ideas discussed during their training, either within their own practice or in community with other professionals. Several authors talk about the value of the role of communities of practice, networking, and peer interaction in supporting professional learning (Simons & Ruijters, 2004; van Schalkwyk et al., 2012; Xu & Brown, 2016), and the importance of situated learning in facilitating transfer (Lave & Wenger, 1991).

To prepare teachers to meet these assessment literacy standards, teacher education programmes may utilize a variety of approaches, including explicit, integrated, and blended assessment education (DeLuca & Klinger, 2010). However, university training programmes must meet several criteria in order to produce assessment literate teachers. From the literature previously discussed, we propose three quality criteria for analysing university training programmes on educational assessment:

Knowledge base: Teacher training courses on assessment must be wide-ranging and cover as many of the important topics as possible. As has been stated, an appropriate knowledge base is necessary for teachers to develop standards and criteria regarding their own assessment practices (Fulcher, 2012; Xu & Brown, 2016).

Time for integration: If teachers are to make significant changes in their assessment practices, it is necessary for them to integrate the knowledge base into their own decision-making processes. This means that the courses must have a sufficiently long duration so that the teachers can reflect on the topics. Schön (1983) has already established the importance of teaching as a reflective profession. It is now broadly understood that teacher learning takes place over time rather than in isolated moments, and that active learning requires opportunities to link previous knowledge with new understanding (Cochran-Smith & Lyttle, 1999).

Practical opportunities: Even if there are general theories about assessment that can be taught in assessment-related courses, in the end, every teacher must translate those theories into their own specific context (Abell & Siegel, 2011). Teacher training programmes must be aware of this and should provide teachers with an opportunity to put their learning into practice in a real context with supervision from their instructors or peers.

Context and research questions

This study is based on Spanish higher education. The pre-service training of Spanish university teachers is often considered insufficient (Sánchez-Moreno & Mayor-Ruiz, 2006), and many university teachers begin their work as teachers with little or no knowledge about assessment. Generally, universities are responsible for designing training programmes, and they also establish the requirements, if any, for teachers to receive positive evaluations.

There are studies that show that Spanish teachers are at a disadvantage when compared to other European teachers (Quesada-Pallarés et al., 2017). In addition, Spanish teachers state that they have difficulty in designing effective and efficient assessment

methods (Fernández Ruiz et al., 2021). Given this, it is extremely important to ensure that university teachers receive high-quality in-service training, and to identify areas for improvement for the future. We explore four research questions (RQs):

RQ1. What are the characteristics of the training courses on assessment?

RQ2. What is the knowledge base covered by the training courses on assessment?

RQ3. How are the scientific foundations on assessment literacy reflected in the teacher training courses?

RQ4. How many university programmes provide high-quality teacher training in assessment?

Methods

Sample

We aimed to analyse all the assessment courses offered for teacher training in all Spanish public universities. Thus, we screened all the teacher training course information from the 50 Spanish public universities. We obtained a total of 1627 training courses. After reading all the course titles to detect common terminologies, those that contained the words "assessment" OR "feedback" OR "examination" OR "test" were selected for further analysis. This produced 110 courses from 31 universities. Unfortunately, as not every course had complete information available, the final sample included a total of 82 courses from 25 universities.

Data analysis

Information about all these 82 courses was downloaded and analysed using the course syllabuses. A syllabus is an explanatory document about a course and is available from most universities. It includes data about the characteristics and contents of the

course. To answer RQ1, data regarding course duration, available places, teaching mode (online, face-to-face, mixed), and assessment methods for evaluating the teachers' learning were extracted. To answer RQ2 and RQ3, data regarding the content of each course were analysed. From a thematic analysis of the course content, 25 themes were found and were grouped into the following six general areas: assessment methods and tools, competence assessment, assessment theory, design and implementation, assessment in specific contexts, and online assessment. To answer RQ4, the following quality standards four university programmes were used: (1) the university offers its teachers at least 30 hours of assessment-related courses; (2) the university offers at least one course which includes supervised practical design and/or implementation tasks; and (3) the university covers at least five of the six general areas mentioned above.

Results

RQ1. What are the characteristics of the training courses on assessment?

There is great variability in the courses analysed as regards the number of places available and the course duration. The courses offer an average of 154.44 places (SD = 241.96) and have an average duration of 8.89 hours (SD = 9.15). These numbers give an insight of how personalized this training can be. As shown in Figure 5.1, most universities offer around 100 places on assessment-related courses, with a total duration of around 20 hours. However, there are universities that escape this trend, in two different ways. On the one hand, universities such as the UM and to a lesser extent the UCM seem to have opted for the massive webinar format, offering many places on their courses, which nevertheless tend to be short. On the other hand, universities like the UMA offer much longer courses with fewer available places.
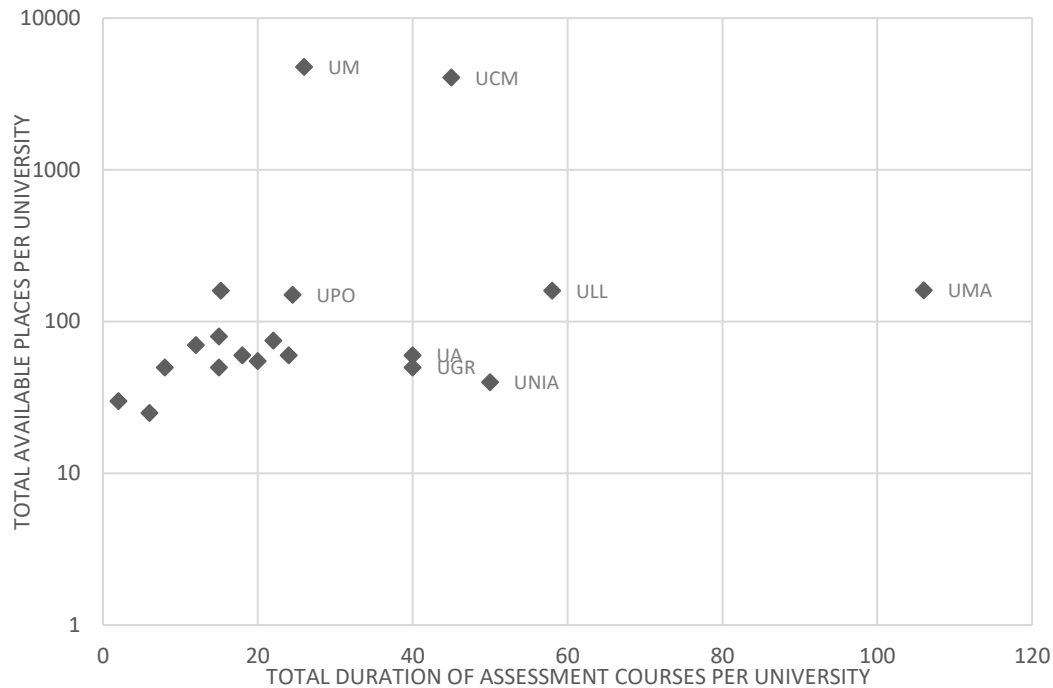
Note: Only universities with complete data available have been included in the figure. To improve readability, Y axis is presented in a log scale.

Figure 5.1. Available places and total duration of assessment-related courses by university

Regarding the teaching modality, 66 of the courses were taught exclusively online, probably as a result of the COVID-19 outbreak during 2020. Another ten courses were exclusively face-to-face, and five courses were mixed. The last course did not offer information regarding its teaching modality.

Lastly, 47 of the courses included an evaluation of the teachers' mastery of their content, and this was usually associated with obtaining a certificate. It should be noted that the same course could use several assessment practices. For 24 of the courses, attendance on the course was used as the evaluation criterion, 15 courses required practical work (e.g. rubric design, Learning Management System workshops), ten courses required assignments, six courses included teaching innovation projects and, finally, five courses had a multiple-choice examination at the end.

RQ2. What is the knowledge base covered by the training courses on assessment?

As is shown in Figure 5.2, a total of 25 themes were identified for the courses analysed, and these were organized into six general areas. We discuss these next.
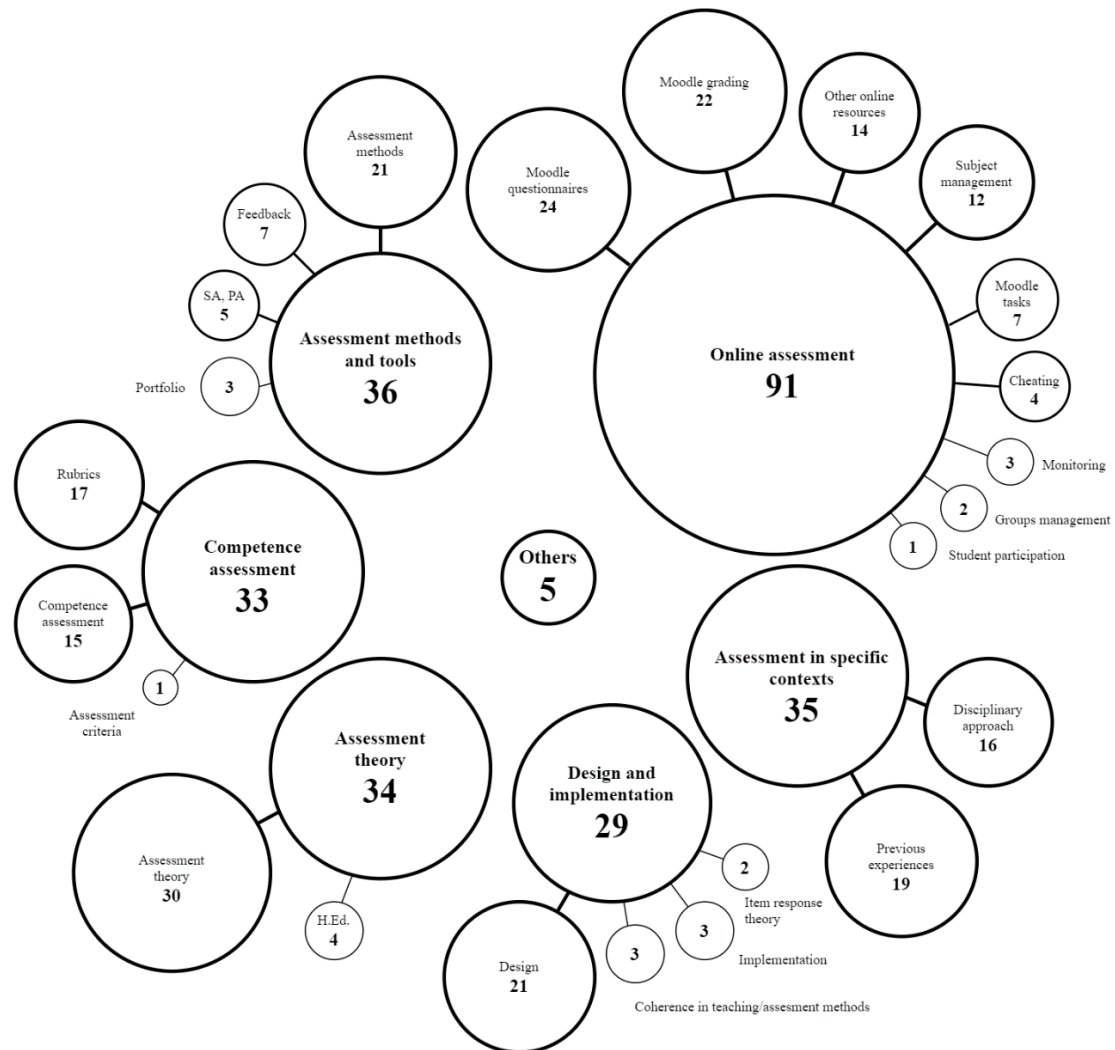

Figure 5.2. Content of assessment-related courses

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

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

A third area is "assessment theory". This area 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 "design and implementation" category includes more practical advice for teachers in carrying out their preferred assessment methods. Several courses address how to design an assessment methodology. Other subtopics in this area are related to the coherence between teaching and assessment methods. Another less common section explains how to achieve an effective implementation of the designed assessment method.

In the category "assessment in specific contexts" we find two large blocks. The first one contains assessment-related courses with a disciplinary perspective (e.g., courses on assessment in health sciences or in STEM disciplines). The second block has to do with the experiences of colleagues in the profession and examples of good practice.

The last category is "online assessment". The strong presence of this topic is probably due in large part to the emergency situation arising from the COVID-19 outbreak. Within this category, several of the uses of the LMS are detailed, such as questionnaires, grades, tasks, and subject management. There are also other courses aimed at preventing contract cheating in online teaching or dedicated to presenting online resources to teachers.

Regarding distribution by university, as can be seen in Figure 5.3 there are several universities, such as the USC, USAL and UNIZAR, that focus their assessment training exclusively on online assessment. At another level, we find universities such as the UDC or the UVA that focus only on two or three of the areas. Universities such as the UA or the UAM offer diverse training, covering practically all the areas found.
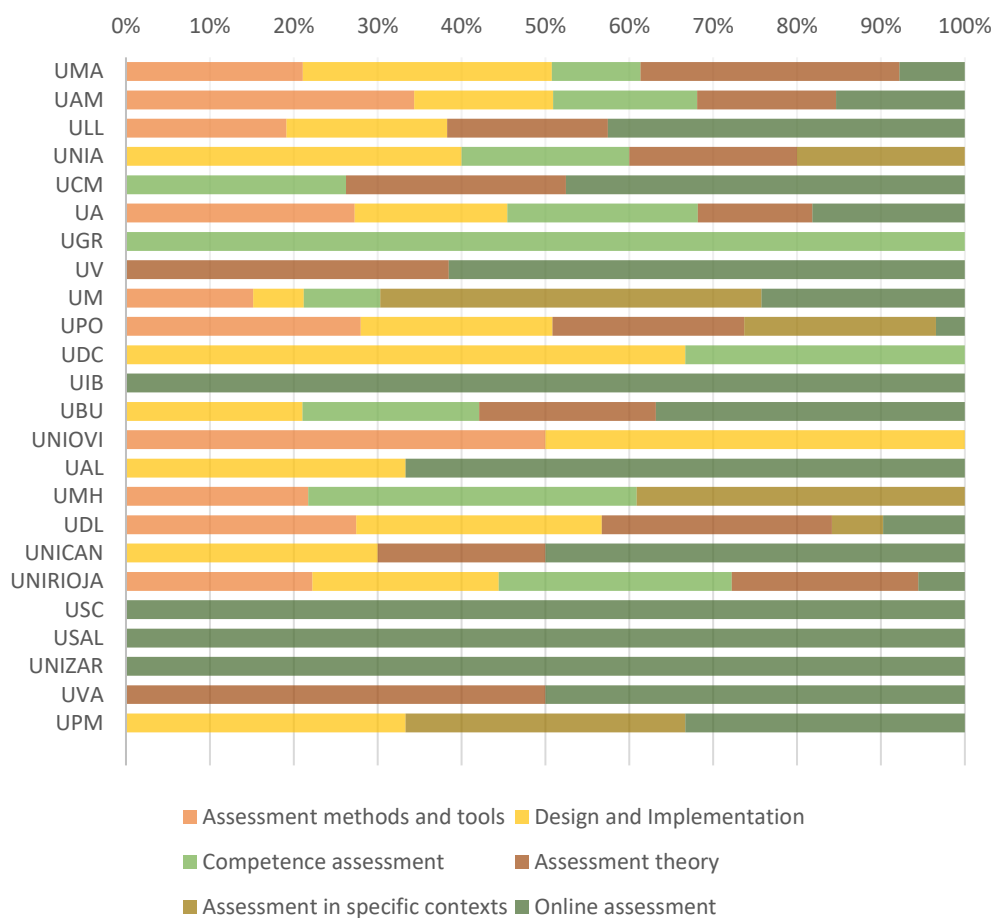
Note: Only universities with complete data available have been included in the figure. Percentage is measured by the total of courses' themes.

Figure 5.3. Content distribution by university

RQ3. How are the scientific foundations on assessment literacy reflected in the teacher training courses?

The previous literature offers recommendations for the characteristics of teacher training courses. Also, the TALiP model includes several knowledge areas that can be

considered as the knowledge base necessary to develop teachers' assessment literacy.

Based on the analysis of the course characteristics and contents, a comparison with literature recommendations and the TALiP knowledge base is presented in Table 1.

Table 5.1

Course characteristics and contents compared with literature recommendations

Literature recommendations	*Course characteristics (Number of courses)*
Assessment education taking various forms and integrating different stakeholders' perspectives (DeLuca, 2012)	*Assessment is usually conceived in an abstract form, barely acknowledging the perspectives of students (4), teachers (1) or faculty environments (10).*
Assessment literacy becoming part of teacher accreditation and certification (Sato et al., 2008)	*Several courses (35) offer a certificate to the teachers, usually after performing some form of assessment of the teachers' progress. However, no university includes mandatory assessment courses for its teachers.*
Mentors attending to student teachers' prior beliefs on assessment (Graham, 2005)	*No explicit data about this. However, the instructor–teachers ratio and the duration of the courses make it difficult to acknowledge individual teachers' beliefs.*
Training content localized and subject-area specific to allow for teachers' free choice (Lam, 2015)	*Several courses (35) are focused on localized situations; this is the case with previous successful experiences with assessment (19) or discipline-specific assessment practices (16).*
TALiP Knowledge Base	*Course contents (Number of courses)*
Disciplinary knowledge and pedagogical content knowledge	*Not included in the analysed courses. Disciplinary knowledge can be assumed because of the teachers' previous training (Masters and PhD programmes). Previous pedagogical knowledge is not a requirement for higher education teachers.*
Knowledge of assessment purposes, contents and methods	*Formative purpose (9) of assessment is covered. A few assessment methods are included, such as portfolios (3) or online questionnaires (24).*
Knowledge of grading	*Several courses (22) address grading in an LMS context, mostly from a technical perspective. Grading in other contexts is not directly addressed.*
Knowledge of feedback	*A few courses (7) address feedback directly. These are mostly oriented towards providing feedback in specific subjects (external practices or theses) or situations (online feedback in COVID times).*
Knowledge of assessment interpretation and communication	*Not very much explored. A few courses addressing feedback (7), rubrics (17), or item response theory (2) can be considered.*
Knowledge of assessment ethics	*There are no courses addressing assessment ethics.*

In summary, most of the areas proposed by the previous literature and the TALiP model are not well represented among the analysed courses. In terms of the course characteristics, there is an absence of different perspectives, and the opportunity to address individual teachers' needs and beliefs in depth is missing. Regarding the content of the courses, a tendency towards treating the teacher as the main element in the assessment processes is observed. Student participation is obviated, either directly (e.g., by the absence of self-assessment or peer assessment in the courses) or indirectly (e.g., by the absence of reflection on the feedback offered to the student or on the justice and ethics of assessment) in most universities.

RQ4. How many university programmes provide high-quality teacher training in assessment?

As reported in the previous section, the criteria chosen to judge whether the assessment training offers "high-quality training" were the following: (1) the university offers its teachers at least 30 hours of assessment-related courses; (2) the assessment course includes at least one supervised practical design and/or implementation task; and (3) the assessment course covers at least five of the six general areas mentioned above. We will analyse the results for each category.

Table 5.2

University scores for the quality criteria

University	Total offer (h)	Supervised tasks	Course content
UMA	106	5	5
UAM	93	3	5
ULL	58	2	4
UNIA	50	0	4
UCM	45	2	3
UA	40	4	5
UGR	40	1	1
UV	32	1	2
UM	26	0	5
UPO	24.5	1	5
UDC	24	0	2
UBU	22	1	4
UIB	22	3	1
UNIOVI	20	0	2
UAL	18	2	2
UMH	18	1	3
UDL	15.25	1	5
UNICAN	15	2	3
UNIRIOJA	15	0	5
USC	14	1	1
USAL	12	1	1
UNIZAR	8	1	1
UVA	6	1	2
UPM	2	0	3
UPF	N/A	0	1

University offers at least 30 hours of assessment related courses. We found eight universities that meet this criterion, with the UMA and UAM spending by far the most time training their teachers in assessment. At a lower level, we found a block of six universities that offer between 30 and 60 hours of assessment-related training.

Assessment courses include at least one supervised practical design and/or implementation task. In this category, there were 18 universities which met the criterion, the majority including one supervised practical task among the assessment methods for their courses. The UA (4) stands out as the university that offers the highest number of supervised practical tasks in its courses.

Assessment course content covers at least five of the general areas. There was no university that covered all six of the general areas detailed above, but seven universities covered five of them in their training programmes, usually omitting "assessment in specific contexts".

From the analysis, we can see that only three of the 25 Spanish universities analysed meet all three proposed quality criteria. These are the UAM, UMA and UA.

In summary, our results present different trends among the universities regarding the characteristics of the courses offered. While there are some universities that choose MOOC formats, others prefer small-group intensive courses. There are a variety of areas covered in the courses, with strikingly different distributions. Online assessment is the most commonly discussed topic, while several universities focus all their assessment training on how to implement it in their LMS. By contrast, relevant areas such as feedback are vastly underrepresented. Lastly, according to the proposed quality standards, only three of the 25 universities included in the study offer what we consider to be high-quality assessment training to their teachers.

Discussion

The aim of this study was to provide an analytical perspective on the training on assessment received by university teachers. Our first research question explored the format and features of the assessment-related courses. The universities analysed presented very diverse characteristics, both in the number of places offered for training courses and in the duration of the courses. For example, the UMA offers several courses of more than 20 hours to a limited number of teachers (up to 40). Other universities like UM, maybe in response to the COVID-19 outbreak during 2020, have opted to offer a large number of massive webinars.

Our second research question explored the topics covered by the assessment-related courses. Using thematic analysis, six general areas were identified in the training-courses content: assessment methods and tools, competence assessment, assessment theory, design and implementation, assessment in specific contexts, and online assessment. These areas can be divided into 25 more specific topics. There are important absences among the courses explored. The low presence of feedback (included in seven courses) and self-assessment or peer assessment (included in five courses) is especially striking.

Feedback is the biggest influence on students' achievement (Carless & Boud, 2018; Hattie, 2009). From the wide empirical research that has been performed, we know that offering high-quality feedback to students is an extremely complex skill (Hattie & Timperley, 2007), and its processes are usually misunderstood by higher education teachers (Boud & Molloy, 2013; Evans, 2013). Likewise, self-assessment and peer assessment are surprisingly rarely mentioned in the analysed courses. Self-assessment has been shown to have positive effects on self-regulation and self-efficacy (Panadero et al., 2017), and is also related to academic achievement (Yan, 2020). The benefits of peer assessment in students' learning (Dochy et al., 1999) and development (Zundert et al., 2010) are also well-known.

Considering the weight of both concepts in educational research and their strong links with academic achievement, their near absence in the training of higher education teachers is concerning. However, this absence could explain assessment trends among European universities, including poor-quality feedback, little use of self-assessment and peer assessment, and traditional assessment methods (Fraile et al., 2017; Ibarra-Sáiz et al., 2010; Panadero et al., 2019).

On the other hand, some topics are widely represented, such as those related to learning management systems. Online assessment is the most common topic in ten of the 25 the universities analysed, with four of them only offering training related to it. Undoubtedly, this has been product of the COVID-19 outbreak, which has created an urgent need for teachers to adapt to online alternatives for assessment (Raza et al., 2021).

Our third research question analysed the course characteristics and contents using the literature about assessment training as a framework. We found that the course characteristics mostly differ from what is recommended, especially as regards stakeholder representation. Assessment is presented as a mostly abstract concept, without it being linked with specific practices, contexts, or agents. The content of the courses also demonstrate several gaps when compared with the TALiP knowledge base (Xu & Brown, 2016). Assessment is mostly represented as a teacher-centred process, with important absences as regards the students' perspective.

The fourth research question asked which Spanish universities offer quality assessment training to their teachers. To define what we considered to be high-quality training, and based on previous models of assessment literacy (DeLuca, 2012; Xu & Brown, 2016), we established three quality criteria, namely: the university offers its teachers at least 30 hours of assessment-related courses, the assessment course includes at least one supervised practical design and/or implementation task, and the assessment course covers at least five of the six previously mentioned general areas. In the following paragraphs, we conduct a more in-depth analysis of what we found in relation to each of these criteria.

Only eight universities offer more than 30 hours of assessment training to their teachers. The threshold of 30 hours was selected because this is equivalent to a 3 ECTS subject, the usual minimum for a Spanish undergraduate degree. However, it can also be

argued that assessment literacy is a complex skill (Xu & Brown, 2016) that can hardly be acquired in less time. Enough time is needed to understand in depth the concepts put forward, and even more time is needed for teachers to perform several iterations in their assessment practices and carry out a critical analysis of what is working, while receiving supervision. There are several intervention studies that show how much more difficult it is to achieve real changes with a short intervention (Janssen et al., 2019; Stes et al., 2010) than with a more extensive intervention (van Aldereen-Smeets & Walma van der Molen, 2015). It is true, and needs to be acknowledged, that professionals have little time, but in-service training can be provided over the whole course of a professional career, and therefore we consider that a minimum of 30 hours of assessment training is an achievable goal.

On the other hand, it was found that 19 universities do not offer any specific training in assessment. This is in line with the results from previous studies (Fernández Ruiz et al., 2021; Mellati & Khademi, 2018; Xu & Brown, 2017) that state that there is a need to develop training programmes for university teachers to increase their assessment literacy.

Teacher training courses and development of assessment literacy

It is important that assessment literacy goes a step beyond the theoretical foundations and the knowledge base, as argued by several of the models on the topic (DeLuca, 2012; Xu & Brown, 2016). During in-service training, this can be achieved through the bidirectional transfer between theoretical content and practical application. In the courses analysed, the outlook is concerning. The theoretical content shows important absences, and the relationships between theory and practice are not sufficiently explored. Only 15 of the courses involved an assessed practical application, and another six included a teaching innovation project.

It is known that teachers have difficulties in implementing changes in their assessment methods (Fernández Ruiz et al., 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 and supervised space in which to try assessment innovations. It is true that teachers have their own concerns and queries arising from their classroom practices (Brookhart, 2002), which may lead to reflection and changes, but, unfortunately, these opportunities do not seem to be widely available for Spanish university teachers.

The lack of these opportunities would also prevent teachers from reaching the next level in the TALiP structure, in which teacher learning is an impetus to achieve assessment literacy. According to Xu and Brown (2016), "The process of becoming literate assessment is fundamentally a transformative, consciousness-evoking one" (p. 157). There are two ways in which this transformation can occur: (a) reflective practice (Schön, 1983) and (b) participation in community activities (Westheimer, 2008). In-service training is ideal for both these methods. First, it gives the opportunity to carry out several trial-and-error iterations, understanding the benefits of each assessment method and testing its results in a supervised and controlled way. Second, taking part in group activities gives the benefit of the exchange of information and experiences among various teachers. We have already seen how the first way is very limited in the current format of in-service training related to assessment. Next, we consider the second way.

There are two methods in the analysed courses that can create community learning. The first consists of supervised group tasks in which teachers can exchange ideas, strategies, and approaches related to assessment. The results of our study show that this practice is rare. Although many courses include homework or practical assignments, these are mostly carried out individually.

The other way to promote community learning would be to design courses whose main purpose is to promote the exchange of ideas among peers. In the courses analysed, this trend can be particularly seen in courses under the theme "assessment in specific contexts". In these courses, either through a disciplinary view of assessment or through sharing experiences of success in the design or implementation of methodologies, teachers are recommended to share their ideas with their peers. However, these courses are very unevenly distributed among the universities analysed, with only a few, such as UM and UMH, covering most of them.

For this reason, it is difficult for the next and higher steps of the TALiP model to be properly achieved. It is hard for teachers to carry out meaningful learning, and, ultimately, for them to acquire an identity as an assessor. It is necessary to emphasize that formal in-service training is not necessarily the only way for university teachers to learn about assessment. Informal learning is recognized as a complementary source of teachers' professional development, but its outcomes are not always comparable to those achieved in systematically supported learning (Hoekstra & Korthagen, 2010). Additionally, vast resources are spent on the design and implementation of these teacher training plans, which makes it especially necessary to ensure their quality.

Conclusion

This study presents a nationwide analysis of in-service teacher training programmes on educational assessment. When contrasted with the existing theoretical models on how teachers achieve assessment literacy, it can be seen that in most Spanish universities these programmes have salient flaws. Important gaps were found in the content of the courses, their duration is too short, and there is a lack of opportunities for the teachers to modify their assessment practices in a collaborative and supervised

environment. Our results make it clear that there is a need for a systematic effort to develop assessment literacy among higher education teachers.

References

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In *The professional knowledge base of science teaching* (pp. 205-221). Springer, Dordrecht.

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.

Airasian, P. W. & Russell, M. K. (2008). *Classroom assessment (6th ed.)*. McGraw Hill, New York.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association (AFT, NCME, & NEA) (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice, 9*(4), 30-32.

Anderson, G. (2006). Carving out time and space in the managerial university, *Journal of Organisational Change Management*, 19(5), 578-592. DOI 110.1108/09534810610696698.

Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart?. *Assessment in Education: Principles, Policy & Practice*, *24*(3), 317-350.

Barnes, N., Fives, H., & EY, C. M. D. (2014). Teachers' beliefs about assessment. In *International handbook of research on teachers' beliefs* (pp. 296-312). 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.

Beziat, T. L., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher*, *27*(1), 25-30.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education: Principles, policy & practice*, *25*(6), 551-575.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, *38*(6), 698-712.

Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, *18*(1), 5-13.

Brookhart, S. M. (2002). What will teachers know about assessment, and how will that improve instruction. In R.W. Lizzitz, W.D. Schafer (Eds.) *Assessment in educational reform: Both means and ends*. Allyn & Bacon, Boston, MA. 2-17.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, *30*(1), 3-12.

Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in education: principles, policy & practice*, *16*(3), 347-363.

Brown, G., Irving, E., & Keegan, P. (2014). *An introduction to educational assessment, measurement and evaluation. Improving the quality of teacher-based assessment*. Dunmore Publishing Limited. Auckland, NZ.

Brown, S. (2014). *Learning, teaching and assessment in higher education: global perspectives*. Macmillan International Higher Education.

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

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325.

Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 1-14.

Chang, S. C., Hsu, T. C., & Jong, M. S. Y. (2020). Integration of the peer assessment approach with a virtual reality design system for learning earth science. *Computers & Education*, *146*, 103758.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational researcher*, *23*(7), 13-20.

Cochran-Smith, M., & Lytle, S. L. (1999). Chapter 8: Relationships of knowledge and practice: Teacher learning in communities. *Review of research in education*, *24*(1), 249-305.

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.

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, *34*(5-6), 576-591.

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, *64*(4), 356-372.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438.

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

Fernández-Ruiz, J., & Panadero, E. (2020). Comparison between conceptions and assessment practices among secondary education teachers: more differences than similarities (Comparación entre concepciones y prácticas de evaluación en profesores de Educación Secundaria: más diferencias que semejanzas). *Journal for the Study of Education and Development*, *43*(2), 309-346.

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

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 471–499). American Psychological Association.

Fostaty Young, S., Young, C. S. F., & Wilson, R. J. (2000). *Assessment and learning: The ICE approach*. Portage & Main Press.

: Fraile, J., Pardo, R., y Panadero, E. (2017). ¿Cómo emplear las rúbricas para implementar una verdadera evaluación formativa? Revista Complutense de Educación, 28 (4), 1321-1334.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132.

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. *Tertiary assessment & higher education student outcomes: Policy, practice & research*, *259*, 275.

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

Hoekstra, A., & Korthagen, F. (2011). Teacher learning in a context of educational change: Informal learning versus systematically supported learning. *Journal of teacher Education*, *62*(1), 76-92.

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

Janssen, E. M., Mainhard, T., Buisman, R. S., Verkoeijen, P. P., Heijltjes, A. E., van Peppen, L. M., & van Gog, T. (2019). Training higher education teachers'

critical thinking and attitudes towards teaching it. *Contemporary Educational Psychology*, *58*, 310-322.

Langford, P. H. (2010). Benchmarking work practices and outcomes in Australian universities using an employee survey. *Journal of Higher Education Policy and Management, 32*(1), 41–53.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Leung, C. (2013). Classroom-Based Assessment Issues for Language Teacher Education. *The companion to language assessment*, *3*, 1510-1519.

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442-467.

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, *20*(5), 523-535.

McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators* (Vol. 1). Corwin Press. Thousand Oaks, California.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational measurement: Issues and practice*, *22*(4), 34-43.

McMillan, J. H. (2017). *Classroom assessment: Principles and practice that enhance student learning and motivation*. Pearson.

Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education (Online)*, *43*(6), 1-18.

Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students. 6th edition.* Pearson Education Inc, Boston.

Observatorio del Sistema Universitario (2021, March 18). ¿A qué puede llamarse universidad? Observatorio del Sistema Universitario.

https://www.observatoriuniversitari.org/es/2021/03/a-que-puede-llamarse-universidad/5/

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., 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.

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.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128-138.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into practice*, *48*(1), 4-11.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265-273.

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*. Ministerio de Educación.

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.

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.*

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, *30*(3), 309-327.

Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice*, *10*(1), 3-6.

Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel recommendations and reality. *Journal of Teacher Education*, *38*(3), 57-63.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Routledge.

Simons, P. R. J., & Ruijters, M. C. (2004). Learning professionals: towards an integrated model. In *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 207-229). Springer, Dordrecht.

Smith, L. F., & Galvin, R. (2014). Toward assessment readiness: An embedded approach in primary initial teacher education. *Assessment Matters*, *7*(3).

Stes, A., Coertjens, L., & Van Petegem, P. (2010). Instructional development for teachers in higher education: Impact on teaching approach. *Higher education*, *60*(2), 187-204.

Stiggins, R. J. (1991a). Assessment literacy. *Phi Delta Kappan*, *72*(7), 534-539.

Stiggins, R. J. (1991b). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, *10*(1), 7-12.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right, using it well*. Assessment Training Institute.

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language testing*, *30*(3), 403-412.

Tierney, R. D. (2012). Fairness in classroom assessment. J. H. McMillan (Ed.) *SAGE Handbook of Research on Classroom Assessment*. 125-144 (Sage).

van Aalderen-Smeets, S. I., & Walma van der Molen, J. H. (2015). Improving primary teachers' attitudes toward science by attitude-focused professional development. *Journal of research in science teaching*, *52*(5), 710-734.

Van Schalkwyk, S., Bezuidenhout, J., Burch, V. C., Clarke, M., Conradie, H., Van Heerden, B., & De Villiers, M. (2012). Developing an educational research framework for evaluating rural training of health professionals: A case for innovation. *Medical Teacher*, *34*(12), 1064-1069.

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.

Westheimer, J. (2008). 41 Learning among colleagues. In M. Cochran-Smith, S. Feiman-Nemser, D.J. McIntyre (Eds.) *Handbook of research on teacher education*, Routledge, New York, NY, pp. 756-783.

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

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241-256.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, *6*(1), 133-158.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *45*(2), 224-238.

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

CHAPTER 6

Discussion

This thesis investigated how university teachers design their assessment methods and the variables influencing them during this process. To do so, three studies were conducted. The first study explored differences in assessment design among academic disciplines and what the differing underlying contextual factors might be. The second study explored how personal factors influenced assessment design, identifying and describing patterns when designing assessment methods. Participants in the first two studies recognised assessment literacy as a major variable in assessment design and claimed to lack the necessary knowledge about assessment. We therefore designed a third study to further investigate the in-service teacher training programmes on assessment nationwide. We compared the characteristics and topics covered with previous literature about assessment literacy to establish quality criteria to help define what makes a high-quality programme for assessment literacy. In the next section, we discuss the findings of each study.

A Summary of Our Findings

Assessment from a disciplinary approach: Design and implementation in three undergraduate programmes. The first study explored the differences in assessment design across three undergraduate programmes representing different disciplines. We found notable differences in various aspects of assessment design: the main purpose of assessment, the preferred assessment methods and the offered for their assessment methods. A summary of the main findings is included in table 6.1.

Table 6.1

Summary of findings: Study 1

	Sport Science	Medicine	Mathematics
Main assessment purpose	Formative	Mixed	Summative
Preferred assessment methods	Continuous assessment	Examinations, simulation tasks	Final and partial examinations
Justification	Learners' professional and personal development	Focused professional context	Diffuse professional context

Regarding assessment purpose, it should be noted that the three degrees offer both formative and summative components in their assessment methods due, among other things, to institutional requirements (Panadero et al., 2019). However, Sport Science teachers tend to design an assessment methodology with a stronger formative approach, in line with the results of Hay (2006), López-Pastor et al. (2013) and Barba-Martín et al. (2020). In medicine, mixed approach is found, as the usual assessment methods include a remarkable accountability approach (Govaerts, 2008), with a constant presence of professional simulation activities as a formative counterpoint (Schuwirtz & Van der Vleuten, 2019). This combination aims to prepare students for two future challenges: the exam for access to the Spanish public health system (MIR) and their performance as medical professionals. The MIR is a high stakes exam that is the entrance to the public national health system; it is used to evaluate the success of the training delivered at medical degrees across universities. Our participants reported a need to design exams similar to the MIR throughout the course, with the aim of preparing students for it. The pressures to use examinations come from both the administration and the students themselves, who are aware of the need to train in situations of high stakes examinations. Lastly, teachers in the mathematics degree present assessment methods with a more summative flavour, at least with regard to the choice of assessment evidence. Closed and

open-book examinations are by far the most used evidence, either with final or mid-term exams. In recent years, that trend is changing in an attempt to divide the contents and evaluate each part separately in a more continuous way. However, the characteristics of this assessment, such as the limited feedback offered, quite dilute this possible formative impulse.

The most striking finding in this study relates to the reasons that teachers offered to justify their assessment methods. Variables that have already been studied in relation to assessment design are cited, such as administrative constraints (Norton et al., 2005; Meyer et al., 2010), exemplified by the use of multiple-choice exams in medicine, or the partial examinations in mathematics. On the other hand, the characteristics of the students are cited as a reason for taking the assessment methods further, when the cohort is especially talented. Finally, workload and infrastructure are cited as the greatest limitations to carrying out the assessment methods that teachers would otherwise choose.

Additionally, the most relevant factor to discuss, given its importance among the participants and the scarce attention it has received in the literature, is the professional context. The division between medicine and mathematics is especially noteworthy. In the case of medicine, both the professional future of the students, as well as their access to it, is perfectly defined. This makes it easy for teachers to design assessment methods that attempt to emulate these challenges, with the aim of preparing students to overcome them. In the case of mathematics, there is such a range of possible professions for its students that teachers cannot predict the skills they will need in their future as professionals. For this reason, its assessment methods are much more focused on content, and leave it up to the student to choose a specific set of skills that they want to promote in further educational stages.

The differences found in the disciplines show how the disciplinary and professional context shape the assessment methodologies even more than the academic characteristics of the degree itself, as other authors have previously stated (Bearman et al., 2017; Carless, 2015; Panadero et al., 2019). After having a look at the external factors affecting the assessment design, the second study of this thesis explores the internal factors.

Assessment design decisions in practice: Profile identification in approaches to assessment design. The main aim of this study was to explore how teachers make decisions during the design of assessment methods, in addition to investigating whether it was possible to identify different profiles in this design. As can be seen in table 2, three profiles were identified, each one showing a specific pattern in the actions carried out and their order. In addition, teachers in each profile present several demographic variables and also report different needs regarding assessment design and implementation.

Table 6.2
Summary of findings: Study 2

	Classic Profile	Competence Profile	Cohesive Profile
Design process	Assessment evidence and assessment grading	Learning outcomes, assessment evidence and assessment grading	Teaching methods, assessment evidence and assessment grading
Individual characteristics	-	More trained and less experienced	-
Reported needs	Assessment literacy	Institutional resources	-

The first profile, the most common among the participants, was the so-called *classic* profile. These teachers tend to focus on the feasibility and efficiency of their assessment methods, and its design process is particularly straightforward. They tend to directly choose a series of assessment evidence that they consider appropriate, and directly distribute the students' final grade to each one of them. While they rarely make additional considerations, some of them paused to reflect on the connection between their

assessment methods and the learning outcomes of the subject or on ways to promote student participation in assessment.

The second profile, also second in popularity, was the so-called *competence* profile. These teachers begin by carefully analysing the learning outcomes and then building a set of evidence that, in their opinion, is the most appropriate for these learning outcomes. Finally, they divide the students' final grade among the set of evidence designed. Once again, it is rare that they make additional considerations, but the use of rubrics it stands out among them.

The last profile has been called the *cohesive* profile. This profile is qualitatively different from the rest, as the teachers in this profile did not design their assessment methods as a separate practice. For teachers with this profile, instructional and assessment methods are an inseparable, and a separate development does not make sense. Assessment design was thus based on specific instructional methods (e.g. project-based learning), after which, the assessment evidence chosen was a direct consequence. In the last step, they would only need to assign a percentage of the grade to each piece of assessment evidence. Teachers in this profile do not make additional considerations.

The identification of the three profiles is the main finding of the second study, but not the only one. We also found that there are two characteristics that differentiate teachers from the competence profile of the rest: teaching experience and previous training. The teachers with the competence profile are the youngest, and therefore the ones with the least teaching experience in higher education and in the subjects they currently occupy. The trend towards a more formative approach among the youngest teachers has been argued by Quesada-Serra et al. (2016) and seems to be supported by these results. However, teachers in this profile were also the most trained, in terms of in-

service training courses, supporting the importance of pre-service training programmes to develop assessment skills among teachers (Picos & López-Pastor, 2013).

There is another difference among teachers depending on their profile: their needs for assessment design and implementation. The teachers in the classic profile stated that they need more knowledge about the different assessment methods, strategies and processes. It is common to observe how these teachers repeat the common assessment methods in their department or faculty, due to a lack of assessment literacy to propose significant innovations. The predominance of this profile seems in line with results such as those found by Ibarra-Sáiz and Rodríguez-Gómez (2010) or Panadero et al. (2019), which show a traditional approach to university assessment. If most of the teachers do not reflect on formative aspects during their assessment design, or lack the resources to put them in practice, it is to be expected that they perpetuate summative approaches, as has been the tradition until now.

On the other hand, teachers of the competence profile consider themselves better trained in assessment, which seems to be the case according to the number of training courses they have taken. In fact, their assessment methods tend to be more developed, and they specify the purpose of each one of them more confidently. However, they have trouble putting their methods into practice. In this sense, they report difficulties in managing the workload created by many of the methods they design. Due to this, in practice they tend to simplify their designs to achieve more efficient methods, even knowing that such concessions are not optimal from the formative point of view.

The two first studies in this thesis present many new findings regarding teachers' actions and needs regarding assessment design. Among the needs, a lack of assessment literacy was constantly present in the first and especially the second study. We then

decided to have an empirical look at the training offered to teachers regarding assessment, which represents the third and last study of this thesis.

A nationwide analysis of in-service courses about assessment for higher education teachers. This study explored the characteristics (RQ1), contents (RQ2) and quality (RQ3) of teacher training programmes on assessment. Regarding the first question, the training courses studied presented very diverse characteristics, both in the number of places offered and in their duration. For example, some universities offer several long courses to a limited number of teachers, while other universities offer a large number of massive webinars.

In relation to the second question, six general areas were identified in the training-course contents: assessment methods and tools, competence assessment, assessment theory, design and implementation, assessment in specific contexts and online assessment. These areas were further divided into 25 more specific topics. The most popular contents were those related to online assessment, possibly due to the pandemic outbreak and the resulting compulsory online education. On the other hand, there are important absences among the courses explored. The low presence of feedback and self- or peer assessment was especially striking.

Regarding the third question, quality criteria were proposed based on theoretical and empirical research on assessment literacy. These criteria were as follows: universities had to offer at least 30 hours of assessment courses, these courses had to have at least one supervised practical activity and they had to cover five of the six main thematic areas described above. However, it was found that only 3 of the 25 universities analysed met all three criteria.

So far we have summarised the main results of the three empirical studies included in this thesis. These studies have been useful for understanding in a much more profound

way how university teachers design their assessment methods, as well as the way in which they are trained for this task. The following sections integrate all of the information obtained in this thesis, first by listing the conditioning variables of assessment design, and second, by proposing a conceptual model for assessment design and implementation.

Variables Shaping Assessment Design

This thesis has been useful for significantly clarifying how assessment design works in the university context. Many of the processes and variables explored in the three studies have been referenced in the previous literature, but there are also variables treated in the studies the existence, or importance, of which has not been acknowledged to date. In the studies, through the testimony of teachers, their decision-making actions and the documents analysed, we investigated how each of these variables influence the way teachers design their assessment methods. Next we present the variables that, according to our results, can be shown to play a role in the way teachers design their assessment methods. These variables are divided into external variables, which were particularly explored in the first study of this thesis, and internal variables, which were more especially explored in the second and third studies. In turn, the external variables are divided into variables related to the discipline and variables related to the environment. In figure 6.1 we observe an overview of all the variables that teachers recognise as influencing their assessment design processes, and in the subsequent sections, we detail how each of these variables influences that process.
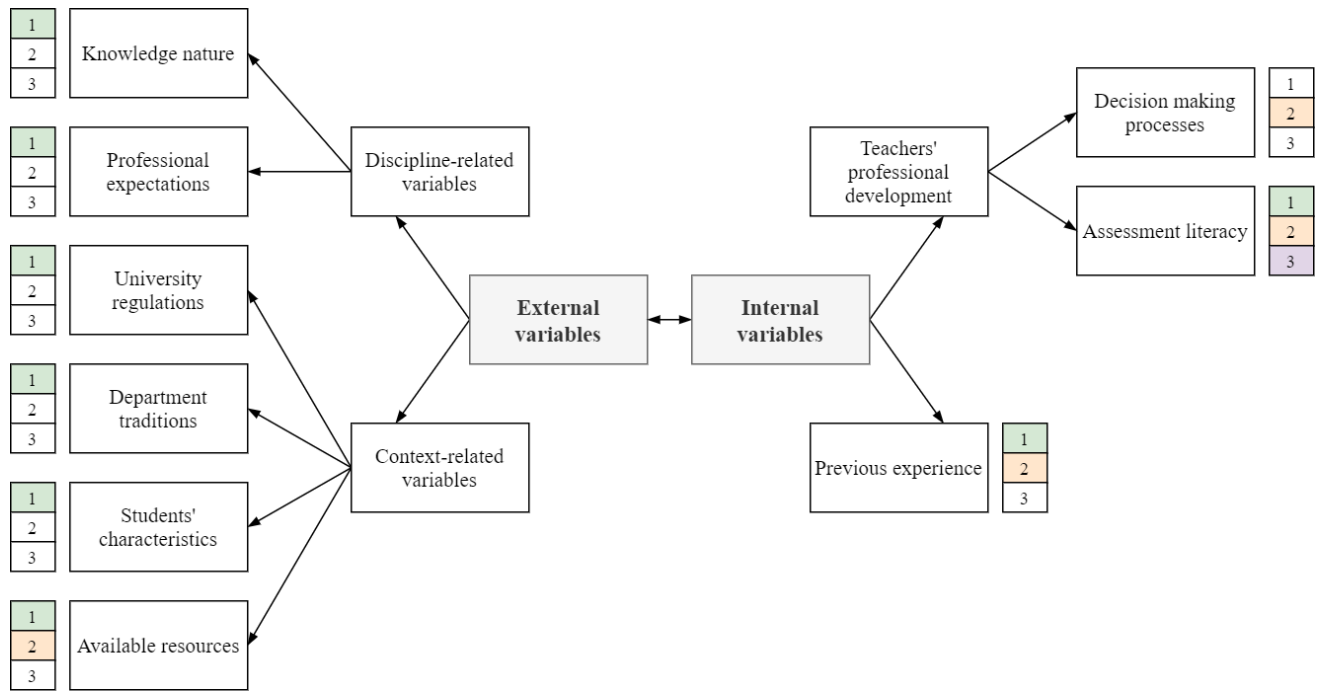
Note: Side numbers represent the studies where each variable was discussed.

Figure 6.1. Variables involved in assessment design (VIAD).

Discipline-Related Variables

The recent literature recognises the situated nature of learning (Jones, 2009; Pryor & Crossuard, 2010), and the influence of the discipline on the assessment practices implemented has been explored in several previous studies (Jessop & Mackelar, 2016; Norton et al., 2005; Panadero et al., 2019). There are differences found in the feedback provided (Esterhazy, 2018; Jessop & Mackelar, 2016) or the presence of authentic assessment (Esterhazy et al., 2021). Regarding teachers' approaches to assessment, it is known that teachers from different disciplines may have different conceptions about assessment (Bearman et al., 2016) or some of its components (Carless, 2015). In our studies, the disciplinary differences regarding assessment design appeared to be due to two major factors: the nature of the knowledge in the field and professional expectations.

Knowledge nature. The teachers declared how certain assessment evidences result more suitable for evaluate specific learning outcomes. Differences may exist within the same discipline (e.g., medicine teachers use a theoretical exam to evaluate a set of

contents, and a simulation task to evaluate practical skills). This is even more apparent when comparing between disciplines. The teachers of the degree in sports sciences or medicine used observation prompts and video feedback (Rucci & Tomporowski, 2010) as an assessment method, both of activities in real time and recorded on video, because the competences to be evaluated can be represented in a video format. This would not be the case in mathematics.

Professional expectations. The influence of professional expectations has been one of the most surprising results of our studies, as it was seldom mentioned in the previous literature. The teachers declare that they prepare their students for their professional future by designing assessment evidence to simulate the intended professional practices of their students. Nevertheless, this is only possible when the career path of the discipline is reasonably clear, as happens for the degrees in medicine. In contrary cases, like mathematics and sport sciences, teachers cannot predict what skills their students will need in the future, and they tend to focus on the content of the subject and the structure of the degree. This has implications that go beyond the differences between disciplines. Even the same discipline in different contexts might have different career paths and therefore different assessment design processes. Lipnevich et al. (2021) carried out a pioneer study comparing assessment methods in Spain and US and found that the distribution by discipline was similar regarding assessment evidence implemented. However, there may still be differences in the characteristics of that evidence.

Context-Related variables

University regulations. Regulations by the university or state institutions (ANECA, in the case of Spain) are also mentioned as a relevant factor in the design of assessment methodologies. Although they are usually generic (Gómez et al., 2013), they

constitute the starting point from which to build specific assessment methods – for example, the obligation of whether or not to offer a final exam. The vagueness of these recommendations is due to the fact that they are made with the intention that each teacher will adapt them to their particular conditions. Therefore, it is assumed that, at a higher institutional level, the recommendations will also be more generic.

Department traditions. As a complement to the university regulations, the teachers of the same department reach a consensus on more specific procedures to carry out in assessment, which would represent the second institutional level. There are certain traditions in departments that function as relevant variables in the design of evaluation methods (Meyer et al., 2010). At this level, teachers hold specific domain knowledge and are more familiar with the characteristics of the subject, which makes it more feasible to reach agreements for modifications and redesigns. Additionally, strong leadership can be present in this layer, especially in small departments (Bearman et al., 2017). This leadership could facilitate innovations and coordination among teachers.

Characteristics of students. . Research about how the individual characteristics of the students mediate their response to assessment is increasing in the literature over the last decade. For example, studies such as Lipnevich et al. (2016) or Jonsson and Panadero (2018) have shown how different students can react to feedback in different ways. However, student characteristics are mentioned as a relevant factor among the participant teachers for different reasons. The first reason is the variability between cohorts, either due to a progressive increase due to demand for the field (e.g. mathematics) or simply due to the differences between one year and the next (e.g. sport sciences). The teachers report that they use more or less ambitious assessment methods depending on the cohort each year. The second reason is student expectations. In degrees like medicine, students are perfectly aware – as teachers are – of the challenges they will have to overcome in

their professional future. This means that they are the ones who demand assessment methods similar to those they will face in the future.

Available resources. This is the most mentioned limiting factor among teachers. Teachers are aware that their assessment methods have flaws, which in their opinion could be solved if they had more resources. According to teachers' testimony, their main need is time to review students' work and provide high-quality feedback. To a lesser extent, several teachers state that the university infrastructure, especially the size or structure of the classes, prevents them from carrying out the assessment methods they would like. Interestingly, studies such as López-Pastor et al. (2013) suggest that, even if the implementation of formative assessment increases teachers' workload, teachers perceive the workload required as much higher than it actually is.

Internal Variables

Decision making processes. Authors such as Bearman et al. (2017) have helped to categorise the decisions that teachers make when design their assessment methods. The profiles found in the second study show how, even without being limited by contextual variables, teachers have different tendencies during this process. The main differences between profiles are how and the order in which they make decisions about which methods to use. While teachers of the classic profile seem to value the feasibility of assessment methods, competence teachers tend to promote alignment of their methods with learning outcomes and cohesive teachers understand assessment as a natural extension of their teaching methods.

Training. Teachers of various profiles detail how a lack of training on assessment limits them when choosing which methods to use in their classes. The demand was so widespread that the third study of the thesis was designed with the aim of analysing teacher training in the field of assessment. Ultimately, teachers need to know the

advantages and disadvantages of the different assessment processes, the variables to consider when designing their methods and the empirical importance of certain assessment strategies and their effects on students. As an example of how training shapes assessment design, there are studies such as Panadero et al. (2016) showing that teachers with more assessment training are more prone to use it in their classrooms. It can also be observed in the second study of this thesis how the teachers who have undergone more training related to assessment are those who are more likely to present a competence profile in the design of their assessment methods.

Experience. As was discovered in the previous literature (Quesada-Serra et al., 2016), experience plays a role in teachers 'attitudes towards innovation and pedagogy. A similar effect is also seen in our studies. Once again, the less experienced teachers tended towards more of a competence assessment design. Additionally, regardless of their general time spent teaching, experience with determined assessment practices works as a predictor for its use in the future. For example, studies by Panadero et al. (2014, 2016) revealed how positive experiences with peer and self-assessment were a strong predictor of their future use.

The variables mentioned above represent a sample of the complexity of assessment design and the quantity of factors that play a role in it. However, the design of assessment methodologies is a sequential process that goes through different areas and actions over time. To integrate the findings of the three studies in this thesis and to obtain a framework of the assessment design and implementation process, we propose a conceptual model of assessment from the teachers' perspective.

An Evidence-Based Model for Assessment Design

Despite the existence of literature about how teachers design their assessment methods, few authors have proposed models that systematically disentangle the processes

in assessment design. Following the results obtained in the three studies of this thesis, and with the purpose of integrating these findings while also offering a necessary contribution to the literature, we propose an evidence-based model for assessment design. The proposal of this model is innovative for two reasons. First, its empirical nature, derived directly from the teachers' observed practices. This differentiates it from previous conceptual contributions such as those proposed by Macdonald and Joughin (2009) or Meyer et al., (2010). Second, it has a sequential nature. This approach to assessment design differentiates this model from the Assessment Design Decisions Framework by Bearman et al., (2016) which offers a series of factors to be taken into account by teachers, but without posing a chronological order.

Our model has two aims, which are coherent with the general purpose of the thesis. The first is to put into context assessment literacy and its importance as a mediator of quality assessment design. The second is to approach assessment design in a systematic way that allows us to understand the different stages of the process and the challenges present in each stage. It must be noted that this represents an open model, which can be adapted to each teacher's individual needs and circumstances. The successive stages can be modified, repeated or skipped in determined contexts.

As can be seen in figure 6.2, in our model, assessment design would be represented as a process with several successive stages. In the following paragraphs we present each of these stages.
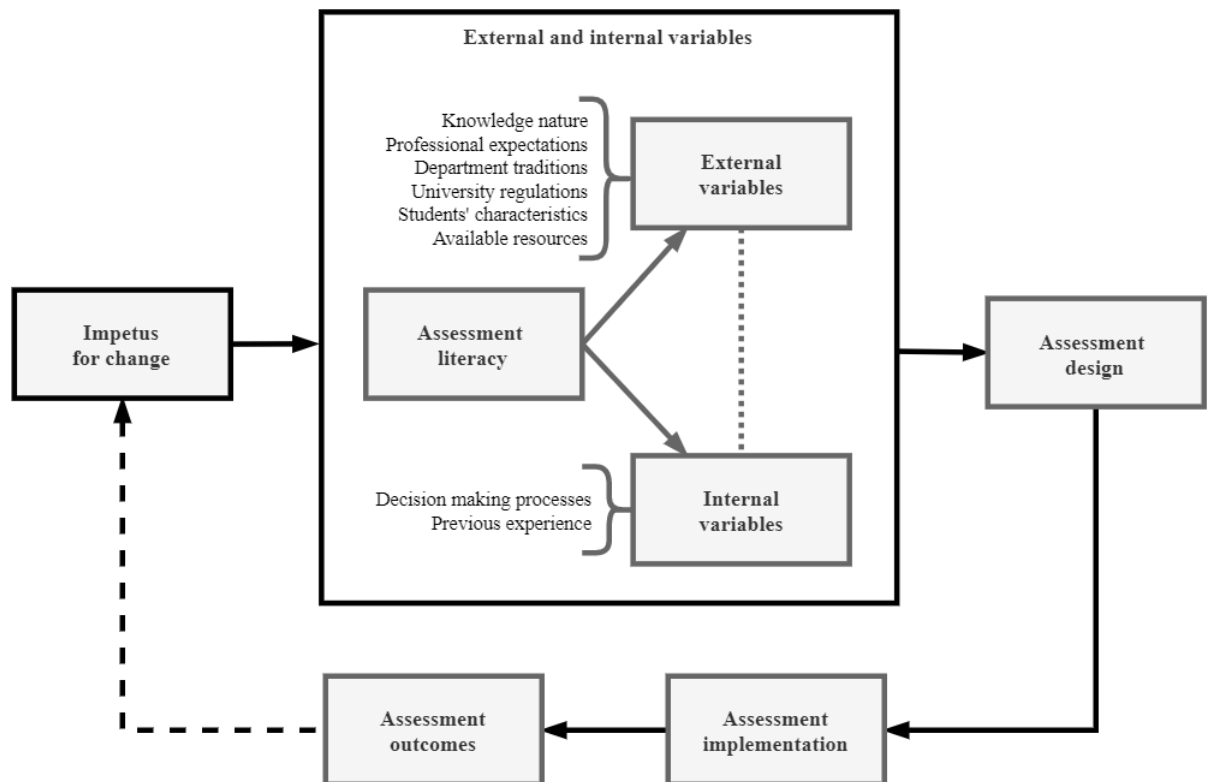
Figure 6.2. Teachers' assessment design and implementation process (TADAIP).

The first phase would be the ***impetus for change*** (Bearman et al., 2017). At this time, teachers would conclude that it is necessary to design assessment methods (either because they are faced with a new subject, new content or a new context), or that it is necessary to redesign the assessment methods that currently exist (either because the results are not adequate or because they want to go a step further in their current methods). It must be noted that the impetus for change is not something that necessarily arises from the teacher him- of herself. At times, external pressures can force teachers to design or redesign assessment methods, even when they consider the task unnecessary.

Once this decision is made, the next step would be where ***assessment literacy*** would come into play. We consider assessment literacy as an internal variable in assessment design, as can be seen in figure 1. However, in this sequential model, it is presented before the rest of the variables as it would come into play at an earlier stage. In

this stage, assessment literacy provides the teacher with a knowledge base of assessment-related methodologies, tools and processes. For teachers to arrive at specific assessment methods, they need to, at least, have been exposed to them. It is enormously difficult for teachers to design an assessment methodology with tools about which they are ignorant. Starting from this, assessment literacy offers a range of possibilities to use. Assessment literacy can be acquired in several ways, both informal and formal. Information training would include the teacher's own experiences, both in as a student and as a teacher, and the knowledge acquired by collaborating with other teachers. Regarding formal training, the most popular option for higher education teachers is both pre-service and in-service training programmes.

This entire range of possibilities is in turn moderated during the next step of the model, which includes the ***external and internal variables***. Some of these variables have been discussed previously. Depending on these variables, whether external or internal, the possibility of the teacher using a specific assessment method increases or decreases. For example, a teacher is more likely to include an exam at the end of the course if all the colleagues in his or her department do it as well. On the other hand, it is unlikely that a teacher will include a portfolio with qualitative feedback if she or he has tried it in the past and it required time to assess that she or he considers excessive.

It is at this point that the ***assessment design*** comes into play. Through a decision-making process, the teacher – starting from more or less concrete ideas, –decides which assessment methods to use (Bearman et al., 2017). As we have seen in this thesis, the design process may be different depending on the teacher. While some start from a set of evidence that seems to be implemented by default, others take time to decide which evidence is most appropriate to evaluate each of the learning outcomes. Finally, and in line with other authors (Brand et al., 2021), some teachers do not understand formative

approaches to assessment as a separate act from instruction. This means that there is no separate assessment design, but rather the instructional and assessment methods of the subject are designed as a single set.

The next step would be ***assessment implementation***. As Carless (2011) argues, this process consists of a series of commitments that teachers must make between their expectations about assessment and the methods they have designed and the professional reality in which they must carry them out. There are many tensions when implementing specific assessment methods, as shown by the fact that teachers design different methods during a fictional task from those they carry out in their classrooms, among many other empirical results (e.g. Norton et al., 2010; Smith, 2011). Teachers are aware that, in practice, they will need to give up many things to be able to implement assessment methods that are as close as possible to what they originally intended.

Finally, once the assessment has been carried out, it would be time to evaluate the ***assessment outcomes***. In these outcomes, there are several variables that the teacher may consider relevant. These variables can come from the students or from the teacher's own perception. Regarding the first, the academic performance of the students can be an indicator of success, especially in formative purposes. Also, the backwash effect, defined as students' response to the assessment methods proposed (Watkins et al., 2005), is considered another outcome of interest for the teacher (Bearman et al., 2017). Regarding the teacher's own perception, his or her general experience, the time and energy consumed and the ease or difficulty with which the methods were implemented could be decisive when reaching a conclusion about their success.

So far we have described the results obtained in this thesis, and we have integrated them into two models, VIAD and TADAIP. In the following sections we discuss several implications, both for professional practice and for future research.

Implications for Professional Practice

The difficulties that teachers face when designing and implementing effective assessment methods form one of the main topics of this thesis. Many of those difficulties are related to the institutional support, including a lack of resources, high workload or difficult coordination between teachers, departments, or faculties. However, one of the most mentioned variables regarding institutional support is the scarce training received in assessment.

Although there are many ways to promote assessment literacy among teachers, the third study of this thesis shows a particularly worrying gap in in-service training programmes on assessment. These programmes are designed, with some exceptions, from a superficial perspective of assessment and have significant lapses in themes and structure. When we compare the characteristics of the courses with the models about assessment design (Xu & Brown, 2016), we can conclude that most of the teacher training programmes would not result in assessment-literate teachers. This thesis has helped to identify the areas of improvement for teacher training programmes in Spanish higher education, and they cover three areas: duration, supervised practice and thematic diversity.

The assessment courses offered at many universities are clearly insufficient. The main example is that there are 19 universities out of 50 that do not offer any course whose main theme is educational assessment. If an in-depth transformation of the knowledge and attitudes that teachers have towards assessment is intended, it will require time. Authors such as Smith and Galvin (2014) defend the importance of giving the teachers time to try new ideas, but time is also necessary for teachers to assimilate new concepts, integrate them into their belief system and put them into practice.

The second area of improvement would be related to this. In-service training is an exceptional opportunity for teachers to carry out controlled processes of trial and error. As previously mentioned, assessment methods implemented in the classroom usually differ from the methods preferred by teachers. This discrepancy between intentions and practice has been addressed by several authors (Norton et al., 2010; Smith, 2011) and may be due to the large number of tensions faced by the teacher (Carless, 2011) and the decisions they make to reconcile those tensions. To identify what these tensions are and how to deal with them in a safe and controlled environment, training courses must offer the opportunity to carry out innovations in the classroom, supervised by the instructor.

Finally, assessment literacy includes a knowledge base (DeLuca, 2012; Xu & Brown, 2016) covering processes, instruments and variables that can be put into play to design assessment methods. It is therefore necessary to expose teachers to the existing trends and variables related to assessment, the empirical evidence for each and their pros and cons, so teachers have the necessary resources to carry out informed decision-making. The courses available should reflect the complexity of a phenomenon such as educational assessment. However, there is a huge lack of courses that talk about specific assessment methods beyond multiple-choice exams. Likewise, many of the processes studied in relation to assessment – such as feedback, student participation or grading – are poorly represented in the courses offered.

The main institutional recommendation extracted from this thesis is the need for assessment training reform, in view of its relevance and its current state. However, with the testimony of the first two studies from this thesis, we can recommend a more systematic approach to higher education teachers for the design of their assessment methods. The effort made by teachers to make explicit the decision-making processes they carry out when designing their assessment methods is noteworthy. This process tends

to be done unconsciously, and teachers tend to have difficulty trying to justify the decisions made (Postareff et al., 2012). Therefore, the proposal of the previously proposed TADAIP model offers a guide for teachers to make explicit the successive steps in the design of their assessment methods. We can expect that if teachers become more aware of the decisions they are making, the sequence in which they are made and the importance and outcomes of each of those decisions, the assessment design and implementation processes can be accomplished more effectively.

As an addition to the implications for professional practice, the questions that have been answered with this thesis give rise to many other questions that need to be explored in future research, which are discussed in the next section.

Implications for Future Research

We present three implications for future research. First, we need to explore how different disciplines differ in the way they design and implement their assessment methods. In addition to the disciplines analysed here – sports sciences, mathematics and medicine – there are other disciplines that could yield interesting results, such as fields in the pure humanities or the arts. Furthermore, according to the importance of career paths in each discipline, as mentioned above, it is necessary to investigate to what extent the results can be replicated within the same discipline in countries with different career paths. Studies such as Lipnevich et al. (2021) have shown the similarities and differences in the assessment evidence preferred in several disciplines from two countries, but a detailed analysis of the characteristics of that evidence is still missing.

The second implication is the replicability of the design profiles found in the second study of this thesis. These profiles, once again, were proposed in a study with a limited number of participants and with an eminently qualitative methodology. The consistency of these profiles needs to be explored in larger samples and from a

quantitative perspective. To do so, the design and validation of self-report methods about assessment design is a necessary starting point. The next step would be a series of subsequent studies to explore the replicability of such profiles. Lastly, studies aiming to identify predictors of each of the profiles (e.g. the internal and external variables of the VIAD model) would produce a powerful contribution to the literature.

The third implication is the importance of the efficiency of the assessment methods. This is a research topic that is scarcely explored in the assessment literature and that, based on the results of this thesis, can be considered of fundamental importance. The scientific literature on assessment has, for decades, been exploring the efficacy of these methods for achieving several variables – with good reason. The effects of numerous assessment instruments, procedures or specific practices have been widely discussed in the scientific research. However, when proposing the implementation of assessment methods supported by scientific research, teachers report finding the efficiency of these methods an important limitation. The literature offers few options for making assessment methods efficient as well as effective. Works such as Higgins et al. (2010) offer a series of initial guidelines. However, it is necessary to develop an empirical and systematic line of research to find out how we can adapt assessment methods so that they consume the smallest amount of resources possible for teachers.

Conclusion

This thesis has allowed us to clarify how university teachers design assessment. The first study explored the external variables that act as conditioning factors in the design of assessment methods, focusing on discipline and paying special attention to career paths, a variable little studied, and which turned out to be of fundamental importance. The second study explored internal variables through an innovative methodology that sought to move away from traditional studies based on self-report. This study allowed us

to identify the existence of individual patterns in assessment design, which can be useful when classifying design actions and establishing intervention paths based on the individual needs of each pattern. Finally, the third study analysed the training offered to university teachers in assessment. This study revealed a series of shortcomings in the planning of teacher training courses, which could help to explain many of the phenomena found in the first two studies and in much of the recent literature.

At this point we can conclude that the design of assessment methodologies is an extremely complex phenomenon that involves numerous variables. However, as found here, it can be accessible for empirical research if we implement some of the data collection and analysis methods used in this thesis. It is therefore necessary to dig deeper into what happens before assessment carried out, but also during and after. It is in the hands of teachers to make the decisions that ultimately are decisive for the learning and development of their students. Although there has been extensive research into understanding the effects that different assessment methods have on students, it is just as important to know the causes behind the use of these methods at the same level. Through this thesis, we contribute as much as possible to explain these causes through the use of innovative methods for the field of assessment.

Conclusión

Esta tesis ha permitido clarificar cómo los profesores universitarios diseñan sus métodos de evaluación. El primer estudio explora las variables externas condicionantes en el diseño de estos métodos, centrándose en la disciplina y prestando especial atención a las trayectorias profesionales, una variable poco estudiada y que ha resultado ser de una importancia fundamental. El segundo estudio explora las variables internas, a través de una metodología innovadora que pretende distanciarse de los estudios tradicionales basados en el autorreporte. Este estudio ha permitido identificar la existencia de

patrones individuales en el diseño de la evaluación, que pueden resultar útiles para clasificar acciones relacionadas con el diseño y establecer procesos de intervención adaptados a las necesidades de cada perfil. Por último, el tercer estudio analiza la formación sobre evaluación que se ofrece a los docentes universitarios. Este estudio revela una serie de carencias en la planificación de los cursos de formación docente, que pueden ayudar a explicar gran parte de los resultados obtenidos en los dos estudios anteriores y en la literatura reciente.

En este punto podemos concluir que el diseño de la evaluación es un fenómeno extremadamente complejo que involucra gran cantidad de variables. Sin embargo, observamos que puede ser accesible para la investigación empírica a través de los métodos de recogida y análisis de datos utilizados en esta tesis. Por tanto, resulta necesario explorar en profundidad qué es lo que sucede antes de que la evaluación sea llevada a cabo, así como lo que sucede durante y después. Está en manos de los docentes tomar decisiones que, en último término, serán determinantes en el desarrollo y aprendizaje de sus alumnos. Aunque existe una extensa literatura para comprender los efectos que los métodos de evaluación tienen en los alumnos, es igualmente importante conocer al mismo nivel las causas tras estos métodos. En esta tesis, contribuimos en la medida de lo posible a explicar estas causas a través de métodos innovadores en el campo de la evaluación educativa.

References

Anderson, L. W. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. Routledge.

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 27*(2), 3-13.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory into practice*, *48*(1), 12-19.

Barba-Martín, R. A., Hortigüela-Alcalá, D., Pérez-Pueyo, Á., & Sánchez-Santamaría, J. (2020). Conceptual Analysis of Influential Factors in the Motivation and Involvement of the University Student towards the Assessment in Physical Education. *Sustainability, 12(21), 8842. doi:10.3390/su12218842.*

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: developing the Assessment Design Decisions Framework. *Teaching in Higher Education*, *21*(5), 545-556.

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.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*:1, 5-25, DOI: 10.1080/0969594X.2010.513678

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5.

Bloom, B.S. (1969). Some theoretical issues relating to educational evaluation. In *Educational evaluation: New roles, new means. The 63rd yearbook of the National Society for the Study of Education, part 2 (Vol. 69)*, ed. R.W. Tyler, 26–50. Chicago, IL: University of Chicago Press.

Boud, D. (2013). *Enhancing learning through self-assessment*. Routledge.

Brand, P. L., Jaarsma, A. D. C., & van der Vleuten, C. P. (2021). Driving lesson or driving test?. *Perspectives on medical education*, *10*(1), 50-56.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, *30*(1), 3-12.

Brown, G. T. L. (2003). *Teachers' conceptions of assessment*. The University of Auckland (New Zealand).

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301-318.

Brown, G. T. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological reports*, *99*(1), 166-170.

Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in education: principles, policy & practice, 16*(3), 347-363.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. Routledge.

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

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in higher education*, *36*(4), 395-407.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The clearing house: A journal of educational strategies, issues and ideas*, *83*(1), 1-6.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*(2), 205-249.

Coffey, M., & Gibbs, G. (2001). The evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK higher education. *Assessment & Evaluation in Higher Education, 26*(1), 89-93.

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, *34*(5-6), 576-591.

Entwistle, N. (1996). *Recent research on student learning and the learning environment.* In Tait, J., & Knight, P. (Eds.). The management of independent learning. Psychology Press, 97-112.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of educational research, 59*(4), 395-430.

Ferguson, P., & T. Womack, S. (1993). The impact of subject matter and on teaching performance. *Journal of teacher education, 44*(1), 55-63.

Fernández-Ruiz, J., & Panadero, E. (2020). Comparison between conceptions and assessment practices among secondary education teachers: more differences than similarities. *Journal for the Study of Education and Development*, 1-38.

Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation, 53*, 69-76.

Fraile, J., Pardo, R., & Panadero, E. (2018). Autoevaluación y autocalificación en el grado de Ciencias de la Actividad Física y del Deporte: estudio censal de las guías docentes. *Revista de Currículum y Formación del profesorado. 22(*3), 164-182.

Gibbs, G. (2006). How assessment frames student learning. In *Innovative assessment in higher education*, ed. C. Bryan and K. Clegg, 23–36. London: Routledge.

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.

Govaerts, M. (2008). Educational competencies or education for professional competence? Medical Education, 42(3), 234- 236.

Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, *21*(6), 607-621.

Guyton, E., & Farokhi, E. (1987). Relationships among academic performance, basic skills, subject matter knowledge, and teaching skills of teacher education graduates. *Journal of Teacher Education, 38*(5), 37-42.

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research papers in education*, *20*(3), 245-270.

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

Hay, P. J. (2006). Assessment for learning in physical education. In *The handbook of physical education,* Edited by: Kirk, D., MacDonald, D. and O′Sullivan, M. 312–325. London: Sage.

Henderson, M., Ryan, T., & Phillips, M. (2019). The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education, 44*(8), 1237-1252

Hernández, R. (2012). Does continuous assessment in higher education support student learning?. *Higher education*, *64*(4), 489-502.

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.

Higgins, R., Hartley, P. & Skelton, A. (2001). Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education 6*(2): 269–74.

Iannone, P., & Simpson, A. (2011). The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications: An International Journal of the IMA, 30*(4), 186-196.

Ibarra-Sáiz, M. S., & Rodríguez-Gomez, G. (2010). An Approach to the Dominant Discourse of Learning Assessment in Higher Education. *Revista De Educación 351*: 385–407.

Knight, P. T. (2002). Summative assessment in higher education: practices in disarray. *Studies in higher Education*, *27*(3), 275-286.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, *32*(2), 169-197.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education, 45*(2), 193-211.

Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied, 15*(4), 319.

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice*, 1-21.

Llamas, M., Mikic, F. A., Caeiro, M., Castro, M., Plaza, I., & Tovar, E. (2018, April). Engineering education in Spain: Seven years with the Bologna Process: First results. In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1775-1780). IEEE.

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*(5), 442-467.

López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2013). Alternative assessment in physical education: a review of international literature. *Sport, Education and Society, 18*(1), 57-76.

Macdonald, R., & Joughin, G. (2009). Changing assessment in higher education: A model in support of institution-wide improvement. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 1–21). Rotterdam, Netherlands: Springer.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational measurement: Issues and practice, 22*(4), 34-43.

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational horizons*, *87*(1), 40-49.

Meyer, L. H., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., & Johnston, P. M. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly, 64*(3), 331-350.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199-218.

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.

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.

Norton, L., Floyd, S., & Norton, B. (2019). Lecturers' views of assessment design, marking and feedback in higher education: a case for professionalisation?. *Assessment & Evaluation in Higher Education*, *44*(8), 1209-1221.

Ödalen, J., Brommesson, D., Erlingsson, G. Ó., Schaffer, J. K., & Fogelgren, M. (2019). Teaching university teachers to become better teachers: the effects of pedagogical training courses at six Swedish universities. *Higher Education Research & Development, 38*(2), 339-353.

O'donovan*, B., Price, M., & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher education*, *9*(3), 325-335.

Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and individual differences*, *22*(6), 806-813.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133-148.

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*(4), 803-830.

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education, 32*(1), 133-156. doi:10.1007/s10212-015-0282-5.

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.

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.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128-138.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press, 2102 Constitutions Avenue, NW, Lockbox 285, Washington, DC.

Picos, A. P., & López-Pastor, V. M. (2013). Haz lo que yo digo pero no lo que yo hago: sistemas de evaluación del alumnado en la formación inicial del profesorado. Do What I Say, Not What I Do: Student Assessment Systems. *Revista de educación*, 361, 279-305.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental?. *Theory into practice*, *48*(1), 4-11.

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265-273.

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.

Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, *38*(3-4), 84-92.

Poulos, A., and Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education 33*(2): 143–54.

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-Serra, V., Rodríguez-Gómez, G., & Ibarra-Sáiz, M. S. (2016). What are we missing? Spanish lecturers' perceptions of their assessment practices. *Innovations in Education and Teaching International*, *53*(1), 48-59.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in higher education, 16*(2), 129-150.

Remesal, A., & Brown, G. T. (2015). Conceptions of assessment when the teaching context and learner population matter: compulsory school versus non-compulsory adult education contexts. *European Journal of Psychology of Education*, *30*(3), 331-347.

Rodríguez Gómez, G., Ibarra Saiz, M., & García Jiménez, E. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de Investigación en Educación, 2 (11), 198-210.*

Rucci, J. A., and P. D. Tomporowski. 2010. "Three Types of Kinematic Feedback and the Execution of the Hang Power Clean*." Journal of Strength and Conditioning Research 24 (3):* 771–778. doi:10.1519/JSC.0b013e3181cbab96

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, *37*(1), 15-24.

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

Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, *45*(3), 669-700.

Schön, D. A. (1984). *The Reflective Practitioner*. New York. Basic Books.

Schuwirth, L. W., & Van Der Vleuten, C. P. (2019). Current assessment in medical education: Programmatic assessment. *Journal of Applied Testing Technology, 20*(S2), 2-10.

Scriven, M. (1967). The methodology of evaluation. In *Perspectives of curriculum evaluation,* ed. R.W. Tyler, R.M. Gagne, and M. Scriven, 39–83. Chicago, IL: Rand McNally.

Simons, P. R. J., & Ruijters, M. C. (2004). Learning professionals: towards an integrated model. In *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 207-229). Springer, Dordrecht.

Smith, J. (2011). Beyond evaluative studies: perceptions of teaching qualifications from probationary lecturers in the UK. *International Journal for Academic Development, 16*(1), 71–81.

Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: the case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education, 38*(1), 44-60.

Stanny, C., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and

    learning through a syllabus review. *Assessment & Evaluation in Higher*

    Education, *40*(7), 898-913.

Tomas, C., & Jessop, T. (2019). Struggling and juggling: a comparison of student assessment

    loads across research and teaching-intensive universities. *Assessment & Evaluation in*

    Higher Education, *44*(1), 1-10.

Trenholm, S., Alcock, L. & Robinson, C. (2015). An Investigation of Assessment and

    Feedback Practices in Fully Asynchronous Online Undergraduate Mathematics Courses.

    International Journal of Mathematical Education in Science & Technology 46(8):

    1197–1221. doi: 10.1080/0020739X.2015.1036946.

Van Schalkwyk, S., Bezuidenhout, J., Burch, V. C., Clarke, M., Conradie, H., Van Heerden, B.,

    & De Villiers, M. (2012). Developing an educational research framework for evaluating

    rural training of health professionals: A case for innovation. *Medical Teacher*, *34*(12),

    1064-1069.

Värlander, S. (2008). The role of students' emotions in formal feedback situations. *Teaching in*

    higher education, *13*(2), 145-156.

Wächter, B. (2004). The Bologna Process: Developments and Prospects. *European Journal of*

    Education 39(3):265–273.

Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of

    assessment: A phenomenographic study of the views of Hong Kong and Swedish

    lecturers. *Instructional Science, 33*(4), 283-309.

Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written

    responses. *Assessment & Evaluation in Higher Education 31(3)*: 379–94.

Wiliam, D., & Thompson, M. (2008). *Integrating assessment with learning: What will it take to*

    make it work?. Routledge.

Winstone, N., & Boud, D. (2019). Exploring cultures of feedback practice: the adoption of learning-focused feedback practices in the UK and Australia. *Higher Education Research & Development, 38*(2), 411-425.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162.

Xu, Y., & Brown, G. T. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, *6*(1), 133-158.

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *45*(2), 224-238.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., ... & Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied measurement in Education*, *21*(4), 335-359.

Young, S. F., Young, C. S. F., & Wilson, R. J. (2000). *Assessment and learning: The ICE approach*. Portage & Main Press.

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