Assessment From a Disciplinary Approach: 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

Recommended citation:

Fernández-Ruiz, J., Panadero, E., & García-Pérez, D. (2021). Assessment from a disciplinary approach: design and implementation in three undergraduate programmes. *Assessment in Education: Principles, Policy & Practice*, 1-21.

Research funded by: Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad) National I+D Call (Convocatoria Excelencia) project reference EDU2016-79714-P; and personal grant (Formación de Personal Investigador) reference BES-2017-080054.

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?

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

S*ample 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.

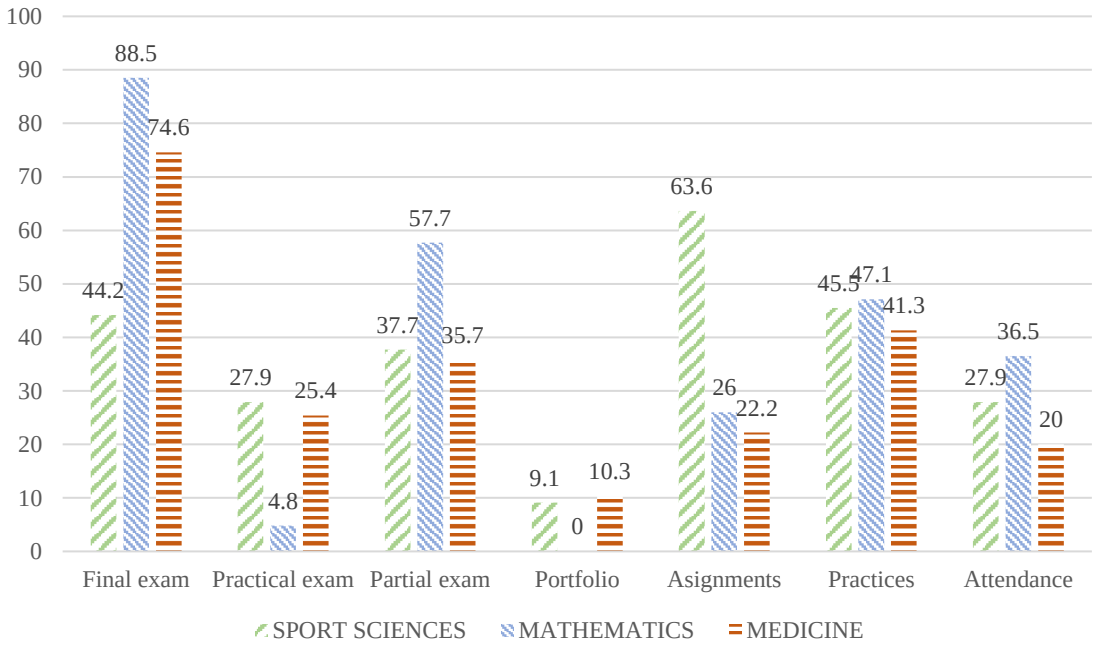
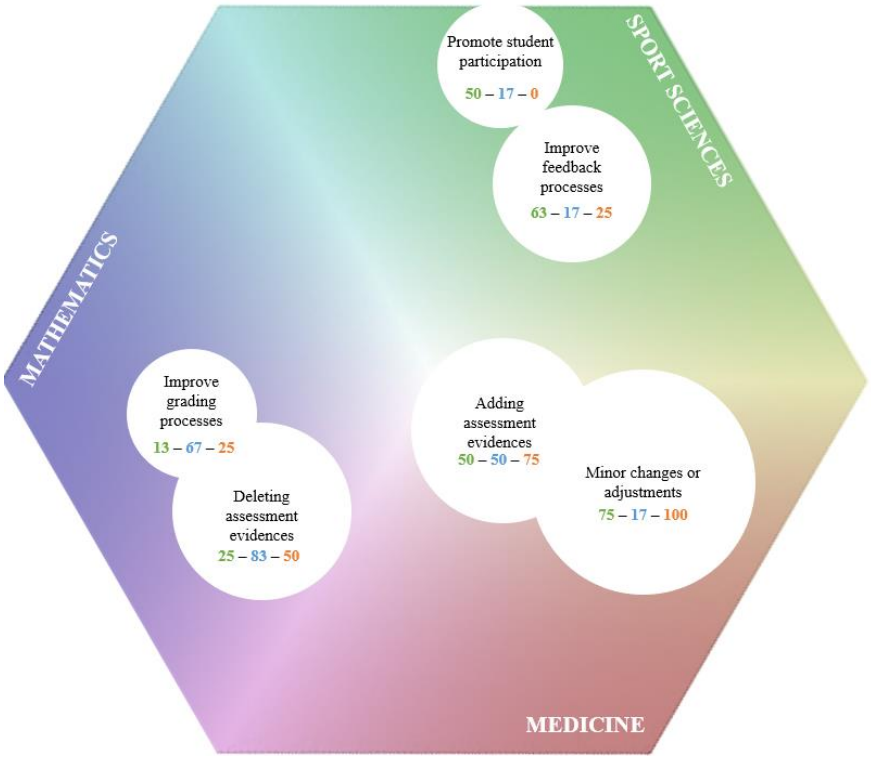


Figure 3.1. Assessment evidences used in each discipline.

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.

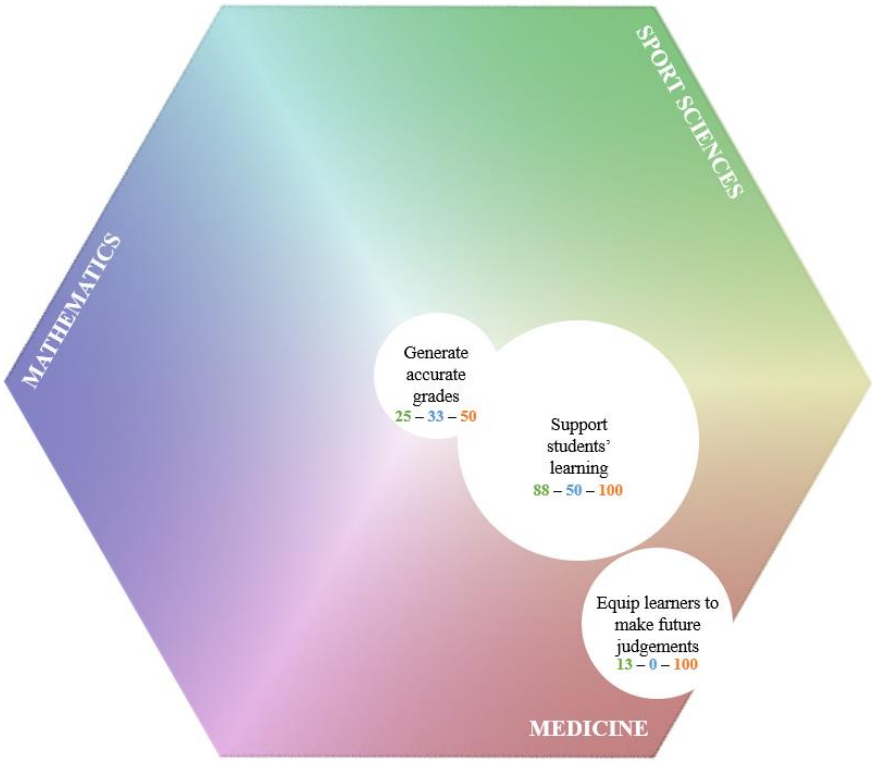
There was not a clear pattern in medicine. Similarly, to those from other disciplines, teachers reported an effort to incorporate new assessment evidence to their subjects. Teachers

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.

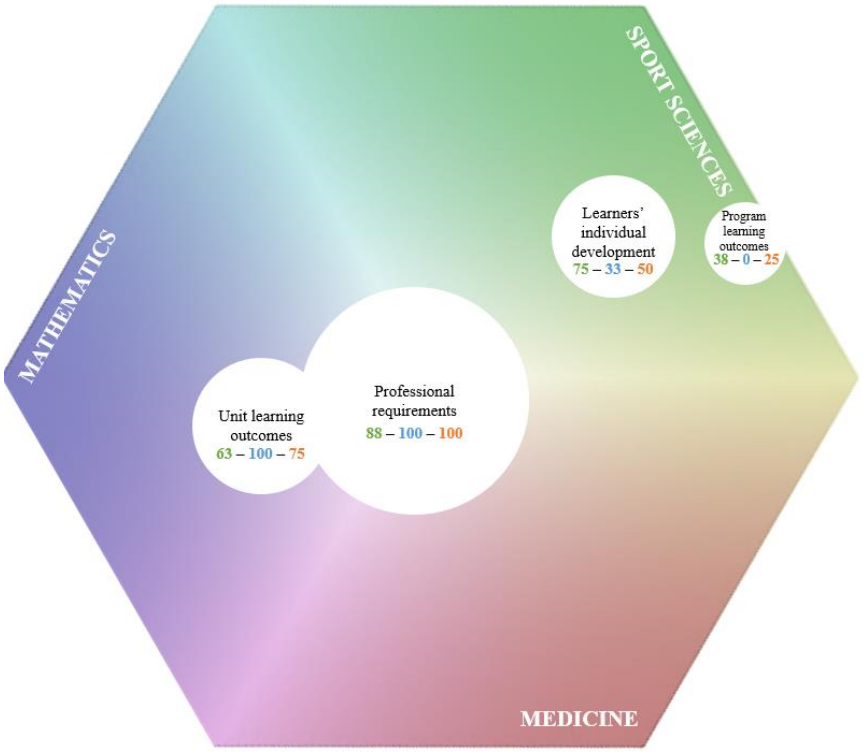
Teachers' answers mostly differed in the formative vision of assessment. Sport science (88%) and medical teachers (100%) were more inclined to understand assessment as a tool for

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.

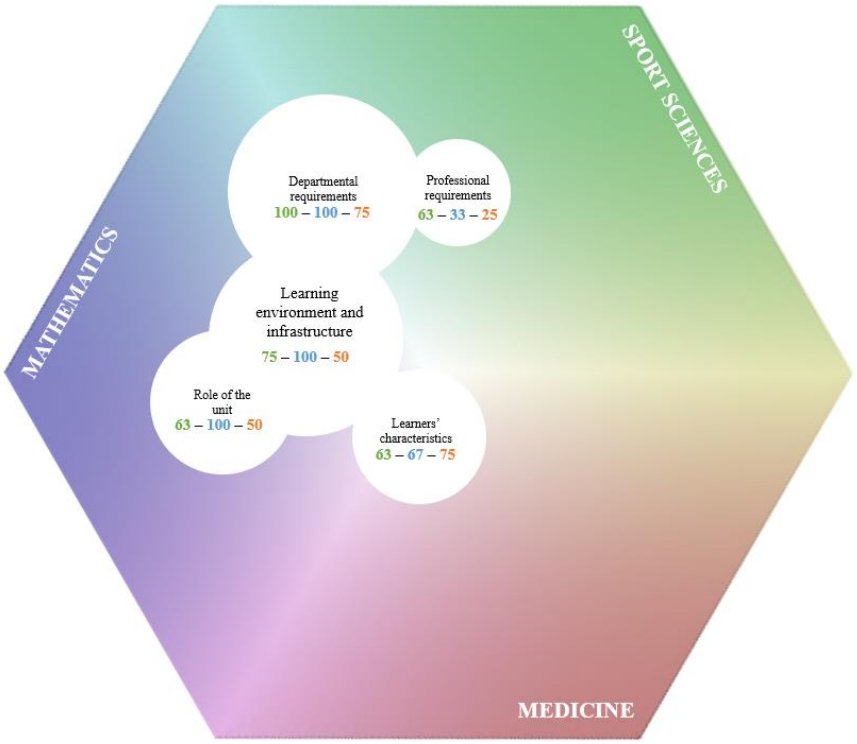
> It is difficult to answer, because in mathematics there are lots of professional possibilities. Depending on what you do, it will be related with my assessment methods or not... But there are people doing very different things, and I cannot guarantee that all I am trying to assess will be applied in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

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

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*

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

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.