Use of formative assessment in Physical Education teacher education in Secondary School: a case study

Uso de la evaluación formativa en formación del profesorado de Educación Física en Educación Secundaria: un estudio de caso

Sonia Asún-Dieste[1] iD

Marta Guíu Carrera[2]

[1] Faculty of Health and Sports Sciences, University of Zaragoza, Zaragoza, Spain

[2] La Marina Secondary School, Santa Cruz de Bezana, Spain

Abstract

The adoption of a competency-based education model in university teaching brings up the debate about the overtaking of traditional assessment methods by new systems, such as formative assessment. The aim of this study was to analyse the level of presence of formative assessment in general and specific courses of the Teacher Education for Secondary School, Vocational Training, Languages, Arts and Sports master's degree, with major in Physical Education, of a Spanish public university, and to explore professors' perceptions. To do so, a complementary mixed quantitative and qualitative design case study was carried out, where data from twelve courses were obtained through the instrument for analysis of syllabus assessment systems (IASEG) and in-depth interviews with five professors. The results revealed the use of formative assessment, with no significant differences between general and specific courses. On the other hand, diversity was detected among professors in terms of their knowledge on this type of assessment and its implementation in the courses. In conclusion, an incipient interest in this type of assessment was observed. However, a number of obstacles were noticed, such as a context that was still unfamiliar with this practice and the lack of training of some professors.

Keywords: assessment, higher education, syllabus.

Resumen

La introducción de un modelo educativo por competencias en la docencia universitaria hace emerger el debate sobre la superación de los métodos de evaluación tradicionales por nuevos sistemas, como lo es la evaluación formativa. El presente estudio tiene por objetivo analizar el nivel de presencia de evaluación formativa en asignaturas generales y específicas de la formación del profesorado de Educación Física del Máster Universitario en Profesorado de Educación Secundaria Obligatoria, Bachillerato, Formación Profesional y Enseñanzas de Idiomas, Artísticas y Deportivas de una universidad pública española y explorar las percepciones del profesorado sobre la misma. Para ello, se realiza un estudio de caso de diseño mixto complementario cuantitativo y cualitativo en el que se obtienen datos de doce asignaturas a partir del Instrumento de análisis de sistemas de evaluación de las guías docentes (IASEG) y de las entrevistas en profundidad realizadas a cinco profesores. Los resultados muestran el uso de evaluación formativa sin que se hayan constatado diferencias significativas entre asignaturas generales y específicas. Por otra parte, se detecta diversidad entre docentes, en cuanto al conocimiento de este modo de evaluación e implementación del mismo en la asignatura. En definitiva, se observa un interés incipiente en este tipo de evaluación; sin embargo, se evidencian obstáculos, como un contexto poco habituado a esta práctica y la falta de formación de algunos docentes.

Palabras clave: evaluación, educación superior, guías docentes.

Formative and Shared Assessment and Feedback: an example of good practice in Physical Education in Pre-service Teacher Education

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
172

Introduction

Assessment is one of the most controversial aspects in the university context. Despite being considered a priority for students and an obligation for professors, there is still no clear consensus on which assessment system is the most appropriate one for university education. In fact, students do not experience a unique assessment model throughout their university studies, but they largely vary depending on the professors who design them, and the centre guidelines and regulations. In any case, according to recent research, it seems evident that traditional assessment models prevail, in which the final exam gains maximum importance (Panadero et al., 2018). Nevertheless, thanks to the change in paradigm at university after the reform towards competency-based university degrees (Álvarez et al., 2014), new assessment models arise that are more in line and consistent with learning than the previous ones, in which student classification and penalty prevailed over their maximum competence development. In this new context, the discussion polarises and professors begin to make decisions regarding the usage of the different models. Undoubtedly, a responsible attitude is needed, which should be based on professors' engagement, as well as on their level of education at that moment (López et al., 2015).

Formative Assessment

Assessment is a process through which information is collected and analysed with the aim to describe reality, make judgements and facilitate decision-making (Sanmartí, 2007). It is one of the most complex tasks within teaching and it allows for understanding and improvement of the practices involved (Santos, 2007). Generally speaking, a lot is tested but very little is assessed during teaching practice, despite students being able to achieve considerable levels of learning through certain assessment processes (Álvarez, 2001; Hamodi et al., 2015). With the purpose to revert this fact, formative assessment has been introduced and developed in different education contexts.

In the university sphere, this model is considered to be an assessment strategy oriented to promoting self-reflection and control over one's own learning. It allows the student to become conscious of how they learn, what they need to do to keep learning and what they need to improve during the learning process (Brown & Pickford, 2013; López-Pastor, 2009; Moraza, 2007). The professor acts as a guide for the student and, thanks to this dialogue process, formative assessment becomes also shared assessment.

Formative and Shared Assessment and Teaching Competence in Higher Education

This type of assessment is now more widely used in higher education after the creation of the European Higher Education Area (EHEA), where exams are only one aspect of the competency acquisition system. Teaching competence is understood as a set of knowledge and capacities that enable an individual to carry out professional teaching and require a formative assessment model (Martínez & Echevarría, 2009). Shared and formative assessment provides a forward-thinking solution to the foundations of the new system established (López-Pastor, 2012). All this is in line with multiple authors who confirmed that it is feasible to teach university courses applying continuous, formative and shared assessment, clearly oriented to student's learning (Busca et al., 2010). Despite it being a newly-emerging topic, there already exist studies showing the benefits of applying formative assessment programmes (e.g. Barberá, 2003; Boud & Falchikow, 2007; Knight, 2005; López-Pastor, 2009). Among them, we can highlight those that associated the implementation of formative assessment systems with better academic performance and higher student engagement (Arribas, 2012; Castejón et al., 2011; Lizandra et al., 2017); those that associated it with improved motivation (Manrique et al., 2012); and those that observed higher student participation (Busca et al., 2010). Potential drawbacks detected were a higher workload for professors and students and a lack of familiarity with this type of assessment systems (Manrique et al., 2012). Additionally, it may happen that professors perceive their own assessment as formative, while students consider it traditional (Gutiérrez et al., 2013).

It must be highlighted that this body of research sometimes used case studies (Hamodi & López, 2012) or action research studies, where formative assessment programmes were applied and learning outcomes, benefits and challenges were analysed (Manrique et al., 2012). Studies using quantitative (e.g. Arribas et al., 2010; Gutiérrez et al., 2013; Hortigüela et al., 2014; Hortigüela et al., 2015; Hortigüela & Pérez-Pueyo, 2016; Palacios & López-Pastor, 2013; Romero-Martín et al., 2017) or qualitative methodologies (e.g. Busca et al., 2010; Hamodi & López, 2012; Martínez & Flores, 2014; Ureña & Ruiz, 2012) are also common. A smaller number of studies applying a mixed methodology have been published. One example is the naturalistic, multi-case study by Castejón et al. (2011) that analysed academic performance and formative assessment in teacher education students. This methodological diversity reflects the complexity of the research topic and responds to the need for understanding the phenomenon from various perspectives.

Formative and Shared Assessment in Physical Education Teacher Education

Assessment system implementation and characteristics become particularly important in Higher Education. It has been generally admitted that an individual tends to reproduce the assessment models they have experienced during their education, perpetuating the same traditional systems and hindering their orientation towards learning and improvement (López-Pastor & Sicilia-Camacho, 2015). According to these authors, the use of traditional assessment systems in pre-service teacher education may have a negative effect beyond university, since students will

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
173

one day become teachers who will very likely reproduce the models received.

Therefore, given its social impact, promoting the implementation of formative assessment systems in teacher education should become a priority. To experience formative and shared assessment as a student has a positive effect on the subsequent teaching performance; however, these experiences are still scarce (Hamodi et al., 2017; Molina & López-Pastor, 2019).

Reality regarding assessment systems in teacher education reveals that there is still a long way to go. A positive evolution has been observed in formative and shared assessment practices in university teaching, but their implementation still needs to be promoted (Romero-Martín et al., 2017) or improved, since sometimes formative assessment systems are used with very limited student engagement (Gallardo-Fuentes et al., 2017). This may be due to the different perceptions of assessment by students and professors.

In the context of Physical Education pre-service teacher education, Martínez, Santos and Castejón (2017) confirmed the difference in perception and discovered that the absence of shared assessment approaches was due to a lack of time, according to professors, and due to a lack of professor's interest, according to students. Romero-Martín et al. (2017) also detected discrepancies between the assessment systems applied and those that were considered as most appropriate in the literature to optimise competency acquisition and learning. This could explain Physical Education teacher education students' limited experience with formative and shared assessment systems (López-Pastor et al., 2016). Nonetheless, the implementation of formative assessment systems in Physical Education teacher education is very positively valued by students given the high transference to their professional career (López-Pastor et al., 2016), and their ability to increase their motivation and learning (Carrere et al., 2017; Hortigüela et al., 2015; Lizandra et al., 2017) and to promote higher-quality and fairer learning, compared to other systems (Atienza et al., 2016).

However, there is a very limited number of studies that have analysed the assessment systems of secondary school Physical Education pre-service teacher education courses through their syllabi. In Spain, where the study was conducted, pre-service teacher education for secondary school is provided in master's degrees divided into several training blocks. The assessment differences among blocks have not been examined in depth. It seems like the general training block, composed of education-related courses, and the specific block, including Physical Education courses, do not apply the same perspective. It is necessary to understand whether professors are still using traditional assessment systems or they are moving towards newly-emerging models. In particular, whether they are paying attention to competencies and focusing on the usefulness of assessment for learning, as formative and shared assessment intends, or not. Therefore, this research consists in a case study in the Physical Education Teacher Education master's degree. It intended to explore the assessment system in order to gain knowledge in this particular context, which could guide the analysis in similar contexts. Besides, it sought to gain evidence in order to better inform future programming decisions in both training blocks.

More specifically, the aim of the present study was twofold: on one hand, to analyse the assessment systems included in the syllabi of the Teacher Education for Secondary School, Vocational Training, Languages, Arts and Sports master's degree, with major in Physical Education, of a Spanish public university, from a formative assessment and teaching competency perspective, distinguishing between general (non Physical Education-specific) and specific courses (directly related to Physical Education); on the other hand, to explore professors' perceptions of the assessment systems applied in this master's degree.

Method

Study Context and Design

The master's degree under study is offered by a Spanish public university. Twenty places were offered in the academic years 2017-2018, 2018-2019 and 2019-2020, and the degree contained 6 general courses, 6 specific courses, an internship and a final year project.

This research consisted in a case study with a complementary mixed-method design, since both quantitative and qualitative elements were included (Cresswell & Plano, 2011). The use of this design was justified because it improves the final outcome through the mutual enrichment of both methods (Ruiz, 2012) and because it is an excellent option to address education research topics (Pereira, 2011).

Sample and Participants

The sample used for the quantitative study was composed of the syllabi of all courses included in the Teacher Education for Secondary School master's degree (with major in Physical Education), except the internship and the final year project, as they used different methods and assessment systems. Twelve syllabi were included in the study.

Of the twelve professors who were giving general and specific courses in this master's degree (8 men and 4 women), five participated in the qualitative study. Mean age was 54 years old (±21), the mean of years of university teaching experience was 21 (±17) and there were 4 men and 1 woman. Three academic ranks were observed: Principal Lecturer, Senior Lecturer and Assistant Lecturer. Three professors had given classes in the master's degree since its creation and two of them had done it for two years. All of them held a PhD.

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
174

Non-probability sampling was applied for participant selection, since the individuals' characteristics were known in advance (Alaminos & Castejón, 2006) and gender and academic rank heterogeneity was prioritised. The number of cases chosen was not a priority; instead, every case's potential was considered the most important aspect to help the researcher (Taylor & Bogdan, 2013).

Instruments

IASEG instrument (instrument for analysis of syllabus assessment systems), validated by the members of the network for formative and shared assessment in education (Red de Evaluación Formativa y Compartida en Educación, REFYCE) (Romero-Martín et al., 2020), was used for the quantitative study. This instrument was composed of 37 items grouped into four dimensions: type of assessment, assessing agent, implicit or explicit feedback and means. The fourth dimension was divided into five sub-dimensions: exams or written tests, sessions, group projects, individual projects and other means. Based on the dimensions and items, an index was obtained to determine the system's competency assessment capacity through formative assessment (ICCD). IASEG instrument, based on syllabus analysis, was used for the quantitative study. First, this tool calculates a value (actual value, AV) for every dimension, which is translated into an ICCD per dimension that can be high, medium or low, depending on the maximum actual value established by the tool. Finally, a global ICCD, ranging between 0 and 100, is obtained per course. The instrument is accompanied by a guide to facilitate the researcher's work (Romero-Martín et al., 2020). This tool allows for analysis of the assessment system applied in a course based on the information available in the university syllabus, applying a formative assessment perspective and measuring its competency assessment capacity.

In the qualitative study, information was collected through in-depth interviews. They were structured into four general observation areas: assessment system, feedback, assessing agents, and means and instruments. Open questions inviting participants to talk were included randomly depending on the interviewees' speech. The interviews were conducted by an expert researcher in qualitative methodology and they were fully recorded by two digital devices.

Procedure

The syllabi of the twelve courses under study, which are publicly accessible, were reviewed in order to collect the data for the quantitative study. Every syllabus was reviewed by two researchers using IASEG instrument in order to increase result reliability. Subsequently, and once the discrepancies had been solved by a third expert, IASEG instrument was used for data analysis, which allowed us to determine whether the elements contained in the instrument were present in the assessment systems or not. Once an index was obtained for every dimension, quartiles were calculated based on ICCD values, and the courses were classified accordingly. The differences among courses were examined based on global and dimension ICCD values.

For the qualitative study, first, the characteristics of the population under study were analysed (gender, age, university teaching experience and academic rank) and professors were contacted to ask for their participation. An informed consent was requested and individual interviews were conducted at the university, lasting for about 40 minutes. To maximise credibility, the interviews were recorded and transcribed; the participants could then read the transcriptions and add the necessary corrections. The data management procedure complied with the guidelines set in the project that comprises this study, which had been previously approved by the Research Ethics Committee with reference number C.P.-C.I.PI21/377.

Data Analysis

Two university experts independently analysed the syllabi and Cohen's kappa coefficient was calculated to determine the level of agreement between them. A high level of agreement was achieved in all syllabus analysis (Kappa=0.92). ICCD was calculated following the procedure proposed by IASEG's authors (Romero-Martín et al., 2020). For the statistical analysis, quartiles were calculated based on course ICCD values. Shapiro-Wilk and Levene tests were conducted to confirm normality and homoscedasticity with a confidence level of 5%. Student's t test for independent samples was applied to compare ICCD mean values for general and specific courses. Nevertheless, normality was not confirmed for some dimensions and sub-dimensions when they were separately analysed. Therefore, Wilcoxon-Mann-Whitney test was applied to examine the differences per dimension, except for Other means, where which Student's t test for independent samples was used. SPSS software (version 23.0) was used for quantitative data analysis.

NVivo 11 software for qualitative research (Sabariego, 2018) was used to analyse the information collected through interviews; a content analysis based on a constructivist approach was conducted (Bardin, 1986). The data were coded and classified into categories (Ruiz, 2012). First, a number (code) was assigned to every interviewee. Secondly, units of information were established and associated with previously existing and emerging data categories. Based on them, a synthesis was proposed and interpreted. Thus, the information collected through interviews not only allowed us to compare and contrast the data with those obtained from syllabus analysis, but it also provided new insights to the research, increasing its credibility and reliability.

Results

The results obtained through ICCD revealed large differences between courses with regard to the use of formative assessment. The most formative course achieved

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
175

74.62 points, while the least formative one obtained 25.78 points, on a 100-point scale.

The assessment system analysis was conducted through quartile calculation.

Based on the quartile values (Table 1), we can state that Q4 (the lowest) contained the three general courses with the lowest ICCD scores (25.78, 26.71, 29.66). Q3 contained two general (30.12, 35.51) and one specific course (33.75). In Q2 there were three specific courses (40.48, 41.02, 41.12). And Q1 (the highest quartile) contained two specific (57.73, 74.62) and one general course (49.73).

Table 1. ICCD quartiles and percentage of courses in every quartile

	Range	Courses (%)
Q1 (highest)	43.27-100	25
Q2	37.99-43.27	25
Q3	30.00-37.39	25
Q4 (lowest)	0-30.00	25

For a deeper analysis, the ICCD values in the instrument's dimensions and sub-dimensions for the twelve courses were examined. The most relevant findings are described below.

With regard to the type of assessment, 33.33% of the courses presented low formative capacity, offering only one global test (ICCD-dimension: type=1.62 points, out of a maximum score of 5.48), while 66.66% of the courses also offered continuous and formative assessment (ICCD-dimension: type=5.48, which is the maximum score).

Regarding the agent dimension, all courses yielded the same value (3.7 points, out of a maximum score of 15.89), having only the professor as assessing agent.

If we look at the feedback dimension, we find that 33.33% of the courses did not include feedback at all (ICCD-dimension: feedback=0), 50% of them presented some implicit feedback in the continuous assessment (ICCD-dimension: feedback=1.8 points, out of a maximum score of 5.78), and 16.66% of the courses showed a high level of feedback in the form of explicit feedback (the ICCD-dimension: feedback reached the maximum score).

Regarding the exams or written tests sub-dimension, in 8.33% of the courses, tests were mostly reproductive in nature and little formative (ICCD-subdimension: exams or written tests=1.8 points, out of a maximum score of 3.88 points), 75% of the courses included tests with greater practical application (their ICCD-subdimension: exams or written tests reached the maximum score of 3.88), and 16.66% of them did not include these means (ICCD-subdimension: exams or written tests=0).

If we analyse the sessions, 75% of the courses did not use this means (ICCD-subdimension: sessions=0), 8.33% of them included session design (ICCD-subdimension: sessions=4.06, out of a maximum score of 15.89), and 16.66% used session design, implementation, observation and assessment/reflection (the ICCD-subdimension: sessions reached the maximum score).

With regard to the group projects sub-dimension, 25% of the courses did not mention this means (ICCD-subdimension: group projects=0), in 33.33% of them, group projects were only theoretical (ICCD-subdimension: group projects=9.65 points, out of a maximum score of 10.92), and 41.66% included theoretical-practical group projects (the ICCD-subdimension: group projects reached the maximum score).

If we examine individual projects, we find that 25% of the courses did not use this means (ICCD-subdimension: individual projects=0), 25% included only theoretical projects (ICCD-subdimension: individual projects=9.29 points, out of a maximum score of 10.56), and 50% of them included practical or theoretical-practical individual projects (the ICCD-subdimension: individual projects reached the maximum score).

In the other means sub-dimension, the course with the lowest actual value did not use any additional means (ICCD-subdimension: other means=0), while the course with the highest actual value used six additional means (ICCD-subdimension: other means=18.21, out of a maximum score of 31.59 points).

When comparing dimensions and sub-dimensions between general and specific courses (Table 2) based on their mean ICCD-dimension values, no differences were found in the agent dimension and slight differences were detected in the exams/tests and sessions sub-dimensions. By contrast, for the rest of dimensions and sub-dimensions, specific courses yielded considerably higher values.

Nevertheless, the statistical analysis did not allow for confirmation of statistically significant differences between general and specific courses, despite the Physical Education-specific courses presenting higher scores than the general ones (Table 2).

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
176

Table 2. Differences in the assessment system by type of course

	Total	General	Specific	pª
Type Mean (±SD)	4.19 (±1.90)	3.55 (±2.11)	4.84 (±1.58)	0.24
Agent Mean (±SD)	3.70 (±0)	3.70 (±0)	3.70 (±0)	.
Feedback Mean (±SD)	1.86 (±2.01)	0.9 (±0.99)	2.83 (±2.39)	0.11
Tests Mean (±SD)	3.06 (±1.55)	2.87 (±1.64)	3.23 (±1.58)	0.59
Sessions Mean (±SD)	2.99 (±6.13)	2.65 (±6.49)	3.33 (±6.37)	0.59
Group projects Mean (±SD)	7.78 (±4.72)	6.86 (±5.34)	8.68 (±4.30)	0.49
Indiv. projects Mean (±SD)	7.60 (±4.62)	5.07 (±5.57)	10.14 (±0.66)	0.11
Other means Mean (±SD)	9.35 (±6.39)	7.31 (±5.43)	11.39 (±7.09)	0.29
ICCD Mean (±SD)	40.52 (±14.30)	32.92 (±8.91)	48.12 (±15.22)	0.06

* Data are ICCD values on a 0-to-100 scale and ICCD values for every dimension.
* Student's t test and Wilcoxon-Mann-Whitney test.

The content analysis conducted in the qualitative study yielded the following categories: training, use, obstacles, agents, instruments and means, feedback and benefits. From all of them we can extract a few ideas that may help understand the situation of formative assessment in the master's degree.

The first one is a large variety in professors' training regarding formative assessment, which generates diversity in meaning interpretation and in-class implementation. Professors with broader knowledge on this topic seem to use it more frequently.

'I have belonged to the network for formative and shared assessment (Red de Evaluación Formativa y Compartida) for a few years. In our yearly gatherings, we discuss our experiences and try to propose new lines and projects. I have broad knowledge on the topic, so I try to reflect carefully on every experience.' (U15, EIAE).

The second one is that the interviewed professors approached assessment from a progress and innovation perspective, trying to avoid traditional practices, such as final reproductive tests, despite acknowledging not having received enough training on new ways of assessment.

'Anything that means innovating, improving, breaking with traditional practices is very important. Anything that helps to generate a learning process that matches social reality is important.' (U12, EIIIAG).

Thirdly, it was confirmed that context and social demands did not guarantee the use of assessment as a learning instrument, but rather these professors tended to set measurement, marking and penalty as the only aims of assessment.

'It is not a disadvantage of formative assessment, but of a context that is not very formative in this regard.' (U14, EIAE).

Fourthly, assessment was observed to be a flexible syllabus element in cases in which these professors included self- and peer-assessment, although they were not mentioned in their syllabi.

'My idea of peer-assessment has many detractors, but I think it is positive to have students participate in their class-mates' assessment [...]. And pretty much the same for self-assessment. In the sense that everyone knows perfectly to what extent they are making progress and how they are achieving it.' (U5, EIAE).

Finally, professors confirmed the existence of advantages and disadvantages of formative assessment system application, since they led to better time management for increased learning, greater student's engagement and continuous learning throughout the course. Nonetheless, they believed that the formative assessment tasks requested to students should be clearer and feedback

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
177

should be provided along the process, instead of at the end of the course.

> 'I think there is no point in assessing the process at the end, if you ask them to make no matter how many reports, but you wait until you have them all before you review them. That's not formative at all.' (U6, EVAE).

They also perceived it could mean a greater effort and higher workload for professors and students.

Discussion

The major study findings revealed differences in the formative nature of the different courses' assessment. However, it cannot be stated that the type of course (general or specific) determined their ICCD, since these differences were not significant.

The data obtained from syllabus design showed that more than half of the courses offered continuous and formative assessment models. This could be because professors tended to apply innovations in the assessment by avoiding more traditional practices like final reproductive tests, using other assessment means and, especially, providing plenty of feedback. Nonetheless, one third of the courses continued to apply the final exam model, the preferred one in the traditional system, while the rest had shifted to a more formative trend.

The study data also revealed differences regarding professors' training on and interpretation of formative assessment, which may have generated discrepancies in the level of application to the different courses. Given the relationship found between training on formative assessment and its application (Molina & López-Pastor, 2019), the lack of training on formative assessment of some professors may have been the reason why means other than the traditional ones, like written reproductive tests, were not included in some of the master's degree courses.

This study confirmed, through syllabus analysis, a limited use of self- and peer-assessment. Despite the fact that evidence in the literature suggests that the participation of other assessing agents other than the professor improves learning processes and outcomes and is positively valued by students (Ibarra et al., 2012), its use is very limited in university education (Arribas et al., 2010). However, in this case, it was detected that self-and peer-assessment could have been used without being specified by professors in their syllabi. Likewise, it was found that not all the information about course assessment instruments was present in these syllabi.

The assessment systems of the syllabi analysed in this study presented a total or partial absence of feedback, which is a clear sign that formative assessment was not sufficiently used to promote learning.

Formative assessment necessarily entails formative feedback, while the use of marks in formative assessment is only meaningful if they are accompanied by the corresponding feedback for the student (Boud & Falchikov, 2007). It is also true that a syllabus may not reflect all the feedback a professor provides in a more spontaneous manner during their teaching practice.

In regard to assessment means, the reproductive nature observed in exams or projects in this study could not be considered as formative. In keeping with this, López-Pastor (2009) stated that the majority of assessment means used at university were rote-learning or problem-solving tests, which were far from professional reality. In fact, only 25% of the courses analysed in this study included session design or implementation in their assessment systems, proving the existence of a gap between professional reality and the formative process.

In the present study, specific courses used a larger number and variety of means, with only a few exceptions. The use of different assessment means that allow for the development of various competencies is a key aspect of teaching. A large diversity of means can be found in the literature, e.g. the study by Hamodi et al. (2015), based on Castejón et al. (2009) and Rodríguez and Ibarra (2011).

In the present study, diversity was observed in formative assessment implementation. The need for continuous training for university professors is an interesting idea. Some studies have addressed this issue and have shed light on the general challenge it is for professors to engage in continuous teaching training. Some major barriers were the difficulty to attend training activities due to schedule incompatibilities or lack of time (Jato et al., 2014) and the resistance to change shown by professors (Bergman, 2014; Hamodi et al., 2017; Zaragoza et al., 2008).

The master's degree professors interviewed in this study described a number of advantages of using formative assessment: better time management for increased learning, higher student participation and engagement —as confirmed in the study by López-Pastor (2009)— and better concept comprehension throughout the process, as evidenced in various studies (e.g. Barberá, 2003; Boud & Falchikow, 2007; Knight, 2005; Ureña & Ruiz, 2012; Zaragoza et al., 2008).

Conclusions

The design of the master's degree courses' syllabi showed heterogeneity as regards formative assessment, with no significant differences between general and specific courses. Nevertheless, in-class assessment could have stronger formative nature than what was observed in syllabus design, as described by professors, who described a spontaneous use of peer- and self-assessment.

The study revealed interesting aspects to be examined in future research, like the fact that the type of master's degree course may have less influence on the use and interpretation of formative assessment than the professor's

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
178

training, social demands, beliefs or the context. Actually, professors acknowledged the benefits of using this type of assessment, which seems contradictory and interesting as a new research line.

One limitation of this study is that it was based on syllabus analysis, but these may not be completely true to life, as professors may adapt or broaden them as needed. A second limitation is that not all professors of this master's degree were interviewed, and there was a large diversity in professor's behaviour regarding assessment. And finally, sample size may have been too small in order to find differences in the use of formative assessment depending on the master's degree course type.

To conclude, it must be highlighted that, given the strong impact that the use of formative assessment may have on future Physical Education teachers and students, it is deemed necessary to conduct more in-depth analyses of the assessment systems that are currently used in all Physical Education teacher education for secondary school degrees, both in syllabi and reality, collecting as many data and involving as many participants as possible.

Funding

References

Alaminos, A., & Castejón, J. (2006). *Elaboración, análisis e interpretación de encuestas, cuestionarios y escalas de opinión.* Editorial Marfil.

Álvarez, J. M. (2001). *Evaluar para conocer, examinar para excluir*. Ediciones Morata.

Álvarez, A., González, J., Alonso, J., & Arias, J. (2014). Indicadores centinela para el Plan Bolonia. *Revista de Investigación Educativa, 32*(2), 327-338. https://doi.org/10.6018/rie.32.2.171751

Arribas, J.M. (2012). El rendimiento académico en función del sistema de evaluación empleado. *RELIEVE, 18*(1), 1-15. http://www.redalyc.org/articulo.oa?id=91624440003

Arribas, J.M., Carabias, D., & Monreal, I. (2010). La docencia universitaria en la formación inicial del profesorado. El caso de la escuela de magisterio de Segovia. *REIFOP, 13*(3), 27-35.

Atienza, R., Valencia-Peris, A., Martos-García, D., López-Pastor, V., & Devís-Devís, J. (2016). La percepción del alumnado universitario de educación física sobre la evaluación formativa: ventajas, dificultades y satisfacción. *Movimento, 22*(4), 1033-1048. http://www.redalyc.org/articulo.oa?id=115349439002

Barberá, E. (2003) Estado y tendencias de la evaluación en educación superior. *Revista de la Red Estatal de Docencia Universitaria, 3*(2), 47-60.

Bardin, L. (1986). *El análisis de contenido*. Akal.

Bergman, M. (2014). An international experiment with eRubrics: An approach to educational assessment in two courses of the early childhood education degree. *REDU. Revista de Docencia Universitaria, 12*(1), 99-110. https://doi.org/10.4995/redu.2014.6409

Boud, D., & Falchikov, N. (2007). *Rethinking assessment in Higher Education. Learning for the long term.* Routledge.

Brown, S., & Pickforf, R. (2013). *Evaluación de habilidades y competencias en educación superior*. Narcea.

Busca, F., Pintor, P., Martínez, L., & Peire, T. (2010). Sistemas y procedimientos de evaluación formativa en docencia universitaria. *Estudios sobre Educación*, 18, 255-276.

Carrere, L. et al. (2017). Descubriendo el enfoque formativo de la evaluación en un curso de Matemáticas para estudiantes de Bioingeniería. *REDU. Revista de Docencia Universitaria, 15*(1), 325-343. https://doi.org/10.4995/redu.2017.6334

Castejón, J., Capllonch, M., González, N., & López-Pastor, V. (2009). Técnicas e instrumentos de evaluación formativa y compartida para la docencia universitaria. En V. López-Pastor (coord.), *Evaluación formativa y compartida en Educación Superior* (pp. 65-91). Narcea.

Castejón, F., López-Pastor, V., Julián, J., & Zaragoza, J. (2011). Evaluación formativa y rendimiento académico en la Formación inicial del profesorado de Educación Física. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte, 11*(42), 328-346. http://cdeporte.rediris.es/revista/revista42/artevaluacion163.htm

Creswell, J., & Plano, V. (2011). *Designing and conducting mixed methods research*. Sage Publications.

Cook, T., & Reichardt, C. (1986). *Métodos cualitativos y cuantitativos en investigación evaluativa*. Ediciones Morata.

Gallardo-Fuentes, F., López-Pastor, V., & Carter, B. (2017). ¿Hay evaluación formativa y compartida en la formación inicial del profesorado en Chile? Percepción de alumnado, profesorado y egresados de una universidad. *Psychology, Society, & Education, 9*(2), 227-238. https://doi.org/10.25115/psye.v9i2.699

Gutiérrez, C., Pérez-Pueyo, Á., & Pérez-Gutiérrez, M. (2013). Percepciones de profesores, alumnos y egresados sobre los sistemas de evaluación en estudios universitarios de formación del profesorado de educación física. Ágora para la Educación Física y el Deporte, 15(2), 130-151.

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
179

Hamodi, C., & López, A. (2012). La evaluación formativa y compartida en la Formación Inicial del Profesorado desde la perspectiva del alumnado y de los egresados. *Psychology, Society, & Education, 4*(1), 103-116. https://core.ac.uk/download/pdf/143458581.pdf

Hamodi, C., López-Pastor, V., & López-Pastor, A. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida en Educación Superior. *Perfiles Educativos, 37*(174), 146-161. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-26982015000100009&lng=es&nrm=iso

Hamodi, C., López-Pastor, V., & López-Pastor, A. (2017). If I experience formative assessment whilst studying at university, will I put it into practice later as a teacher? Formative and shared assessment in Initial Teacher Education (ITE*). European Journal of Teacher Education, 40*(2), 171-190. https://doi.org/10.1080/02619768.2017.1281909

Hortigüela, D., Abella, V., & Pérez-Pueyo, A. (2014). Perspectiva del alumnado sobre la evaluación tradicional y la evaluación formativa. Contraste de grupos en las mismas asignaturas. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 13*(1), 1-14.

Hortigüela, D., Abella, V., & Pérez-Pueyo, A. (2015). Percepciones del alumnado sobre la evaluación formativa: Contraste de grupos de inicio y final de carrera. *REDU. Revista de Docencia Universitaria, 13*(3), 13-32. https://doi.org/10.4995/redu.2015.5417

Hortigüela, D., & Pérez-Pueyo, A. (2016). Percepción del alumnado de las clases de educación física en relación con otras asignaturas. *Apunts, 123*, 44-52. http://dx.doi.org/10.5672/apunts.2014-0983.es.(2016/1).123.05

Ibarra, M., Rodríguez, G., & Gómez, M. (2012). La evaluación entre iguales: beneficios y estrategias para su práctica en la universidad. *Revista de Educación, 359*, 206-231. http://dx.doi.org/10-4438/1988-592X-RE-2010-359-092

Jato, E., Muñoz, M., & García, B. (2014). Las necesidades formativas del profesorado universitario: un análisis desde el programa de formación docente de la Universidad de Santiago de Compostela. *REDU. Revista de Docencia Universitaria, 12*(4), 203-229. https://doi.org/10.4995/redu.2014.5621

Knight, P. (2005). *El profesorado de Educación Superior*. Narcea.

Lizandra, J., Valencia-Peris, A., Atienza Gago, R., & Martos-García, D. (2017). Itinerarios de evaluación y su relación con el rendimiento académico. *REDU: Revista de Docencia Universitaria, 15*(2), 315-328. https://doi.org/10.4995/redu.2017.7862

López, M., Pérez-García, M., & Rodríguez, M. (2015). Concepciones del profesorado universitario sobre la formación en el marco del espacio europeo de educación superior. *Revista de Investigación Educativa, 33*(1), 179-194. https://doi.org/10.6018/rie.33.1.189811

López-Pastor, V. (2009). *Evaluación formativa y compartida en Educación Superior. Propuestas, técnicas, instrumentos y experiencias*. Narcea.

López-Pastor, V., Pérez-Pueyo, A., Barba, J., & Lorente-Catalán, E. (2016). Percepción del alumnado sobre la utilización de una escala graduada para la autoevaluación y coevaluación de trabajos escritos en la formación inicial del profesorado de educación física (FIPEF). *Cultura, Ciencia y Deporte, 11*(31), 37-50. http://www.redalyc.org/articulo.oa?id=163044427005

López-Pastor, V., & Sicilia-Camacho, A. (2015). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education, 42*(1), 77–97. https://doi.org/10.1080/02602938.2015.1083535

Manrique, J., Vallés, C., & Gea, J. (2012). Resultados generales de la puesta en práctica de 29 casos sobre el desarrollo de sistemas de Evaluación formativa en docencia universitaria. *Psychology, Society & Educación, 4*(1), 87-102.

Martínez, L., Castejón, F., & Santos, M. (2012). Diferentes percepciones sobre evaluación formativa entre profesorado y alumnado en formación inicial en educación física. *Revista Electrónica Interuniversitaria de Formación del Profesorado, 15*(4), 57-67.

Martínez, L., & Flores G. (2014). Profesorado y egresados ante los sistemas de evaluación del alumnado en la formación inicial del maestro en educación infantil. *RIDU, 1*. https://doi.org/10.19083/ridu.8.371

Martínez, L., Santos, L., & Castejón, F. (2017). Percepciones de alumnado y profesorado en Educación Superior sobre la evaluación en formación inicial en educación física. *Retos, 32*, 76-81.

Martínez, P., & Echevarría, B. (2009): Formación basada en competencias. *Revista de Investigación Educativa, 27*(1), 125-147.

Molina, M., & López-Pastor, V. (2019). ¿Evalúo cómo me evaluaron en la Facultad? Transferencia de la evaluación formativa y compartida vivida durante la formación inicial del profesorado a la práctica como docente. *Revista Iberoamericana de Evaluación Educativa, 12,*1, 85-101. https://doi.org/10.15366/riee2019.12.1.005

Moraza, J. (2007). *Utilidad Formativa del Análisis de Errores Cometidos en Exámenes (Tesis doctoral).* Universidad de Burgos, España.

Palacios, A., & López-Pastor, V. (2013). Haz lo que yo digo pero no lo que yo hago: sistemas de evaluación del alumnado en la formación inicial del profesorado.

Use of formative assessment in Physical Education teacher education in Secondary School: a case study
Asún-Dieste et al.
180

Revista de Educación, 361. https://doi.org/10-4438/1988-592X-RE-2011-361-143

Panadero, E., Fraile, J., Ruiz, J.F., Castilla-Estévez, J., & Ruiz, M. A. (2018). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2018.1512553

Pereira, Z. (2011). Los diseños de método mixto en la investigación en educación: Una experiencia concreta*. Revista Electrónica Educare, 15*(1), 15-29. http://www.redalyc.org/articulo.oa?id=194118804003

Rodríguez, G., & Ibarra, M. (2011). *E-Evaluación orientada al e-aprendizaje estratégico en educación superior*. Narcea.

Romero-Martín, R., Castejón-Oliva, F., López-Pastor, V., & Fraile-Aranda, A. (2017). Evaluación Formativa, Competencias Comunicativas y TIC en la formación del profesorado. *Comunicar, 25*(42), 73-82. https://doi.org/10.3916/C52-2017-07

Romero-Martín, R., Asún-Dieste, S., & Chivite-Izco, M. (2020). Diseño y validación de un instrumento para analizar el sistema de evaluación de las guías docentes universitarias en la formación inicial del profesorado (IASEG). *Profesorado, Revista de Currículum y Formación del Profesorado, 24*(2), 346-367. https://doi.org/10.30827/profesorado.v24i2.15040

Ruiz, J. (2012). *Metodología de la Investigación Cualitativa* (5th ed.)*. Publicaciones de la Universidad de Deusto.

Sabariego, M (2018). *Análisis de datos cualitativos a través del programa NVivo 11PRO*. Universidad de Barcelona.

Sanmartí, N. (2007). *10 ideas clave, evaluar para aprender*. GGraó.

Santos, M. A. (2007). *La evaluación como aprendizaje: la flecha en la diana.* Bonum.

Taylor, S., & Bogdan, R. (2013). *Introducción a los métodos cualitativos de investigación*. Paidós.

Ureña, N., & Ruiz, E. (2012). Experiencia de evaluación formativa y compartida en el Máster Universitario en Formación del Profesorado de Educación Secundaria. *Psychology, Society & Education,* 4(1), 29-44.

Zaragoza, J., Luis, J., & Manrique, J. (2008). Experiencias de innovación en docencia universitaria: resultados de la aplicación de sistemas de evaluación formativa. *REDU. Revista de Docencia Universitaria, 4*, 1-33. https://doi.org/10.4995/redu.2009.6232