Impact of a Training Programme on the e-rubric Evaluation of Gamification Resources with Pre-Service Secondary School Science Teachers

Antonio-Joaquín Franco-Mariscal[1] · Daniel Cebrián-Robles[1] ·
Noela Rodríguez-Losada[1]

Abstract

Social constructivism is a learning approach in which students actively construct their own knowledge by way of experiences and interactions with others. As such, it is important to highlight both individual and group-based reflection practices in pre-service teacher training as a key aspect for improving teaching practice. This paper presents the results of the design and implementation of a training programme of 7 sessions (8.5 h of class participation plus 25 h of personal homework) for pre-service secondary school science teachers, who were asked to design a gamification resource and an e-rubric to evaluate it. Subsequent improvement of this e-rubric was enhanced by performing different reflection activities at key moments. The programme was carried out by 50 Spanish pre-service teachers from Málaga (Spain). Data collection centred on the e-rubrics designed, the emotions experienced and the possible transfer to real-life practice performed six months later. The impact of reflections on the evaluation was studied by analysing the evolution of the categories proposed by the participants for the e-rubric at different times, with marked changes being found during design and preparation of the gamification resource, and only very minor changes post-implementation. In addition, a group-based criteria consensus session favoured a more in-depth reflection. Interest was the main emotion experienced by pre-service teachers, especially during preparation and use of the resource. The programme also had a marked impact on transfer of the e-rubric into practice, as did the designed resource, although to a lesser extent.

Keywords Reflective teachers · Pre-service secondary science teachers · Social constructivism · Evaluation · Gamification resources

✉ Antonio-Joaquín Franco-Mariscal
anjoa@uma.es

Daniel Cebrián-Robles
dcebrian@uma.es

Noela Rodríguez-Losada
noela@uma.es

[1] Didáctica de Las Ciencias Experimentales, Facultad de Ciencias de la Educación, Universidad de Málaga, Campus de Teatinos s/n. 29071, Málaga, Spain

 Springer

1 Introduction

The reflection of teachers regarding their practice is essential during pre-service training (Beauchamp, 2015; Smith et al., 2017). Studies published by Theodoulides and Armour (2001) or Slepcevic-Zach and Stock (2018) concerning the reflections performed by teachers regarding their teaching practices during pre-service training highlight that these reflections may serve to self-regulate learning and may also be useful once they enter service. So, when they are in-service teachers, they could also place greater emphasis on helping adolescents to self-regulate their own learning (Sebre & Miltuze, 2021). However, the review by Van Beveren et al. (2018) into reflection activities at university shows the need to carry out further empirical studies that may help to highlight the type of reflection activities that work best. In addition, although reflection practices are encouraged in the classroom, there is often little or no preparation or instructions explaining how to carry them out (Russell, 2005).

Lee (2007) argues that it is very interesting for pre-service teachers to reflect on their teaching practice as soon as possible in order for them to learn how to resolve any difficulties they may come across as regards the difference between their vision of teaching and reality.

This reflection forms part of constructivism, which supports the idea that students themselves should construct their own knowledge as an active participant in their own learning process (Poerksen, 2004; Woolfolk, 2014). One such interesting model is known as social constructivism, where students actively construct their own knowledge by way of experiences and interactions with others (McKinley, 2015; Rannikmäe et al., 2020) and in which group-based reflection is of particular importance. Social constructivism is based around the social nature of cognition and proposes an educational framework that allows students to become involved in discussion and reflection, encouraging them to argue their ideas in order to share them with others and seek autonomy by interacting for mutual benefit (Akpan & Kennedy, 2020). Numerous factors are known to influence learning and their reflection concerning practice as part of this approach, thus affecting both the activities proposed and the evaluation thereof. Social constructivism emphasises the importance of having students actively involved in their learning process (Guillén et al., 2021).

In order to maintain the interest of students in this type of reflexive activities, which are generally associated with a low "emotional climate" (Bellocchi et al., 2014), interesting, active and entertaining proposals, such as those related to gamification, are required. Polin (2018) has highlighted the potential of educational games from a social constructivism point of view as they allow the creation of attractive, authentic, complex and collaborative spaces for reflection and learning for both students and teachers alike. One alternative with significant advantages is the construction of games by and with students. However, this requires an in-depth understanding of the subject and a significant time commitment for the construction and reflection activities in class (Polin, 2018).

Social constructivism is also present in science fairs, when students explain a project they have designed to their companions (Duit & Treagust, 1998). Paul et al. (2016) concluded that communication processes between students in a science fair promote a meaningful reflection about their project.

Lee's study (2007), and the opinions of teachers reported in the studies by Ryan et al. (2017) or Goulette and Swanson (2019), confirms that the evaluation process is one of the most complex tasks in teaching and, as such, we believe that it should receive greater attention from the very earliest stages of training. As such, reflection of the evaluation in science

education from a social constructivism perspective requires alternative evaluation methods based on the students' different learning styles in order to offer them all the opportunity to express themselves (Iofciu et al., 2012). In this regard, electronic rubrics may help and facilitate this reflection on evaluation, and may also aid dialogue and feedback between teacher and student (Lu & Zhang, 2012). Thus, students can interact amongst themselves using digital tools and develop communication and social skills by way of peer evaluation.

One final yet important aspect of reflection concerns the relationship between social learning and emotions. In other words, the close interactions between cognitive, conative and affective factors during students' learning and problem-solving (Op'T Eynde et al., 2006). An understanding of the role of emotions in science education, and the ability to reflect on them, implies an understanding of the nature of the cognitive processes involved.

In light of the above, this study attempts to incorporate educational activities for pre-service secondary science teachers (hereinafter PSTs) into the framework of social constructivism in science education. The originality of this study lies in several aspects:

(1) This paper presents a training programme for PSTs integrating several key constructs of social constructivism (reflection on practice, gamification resources and science fairs, e-rubric assessment and social-emotional learning). This study links all these aspects in science education, which have traditionally been considered individually, together.

(2) The programme involves PSTs in the design, implementation and evaluation of gamified resources using e-rubrics, as well as the emotions experienced during these activities by way of a formative and continuous reflection for experiences.

(3) The programme uses science fairs to present and reflect on gamification resources and their evaluation. Unfortunately, this format has received little attention in initial teacher training in Spain.

(4) This study would research the impact of the programme on transfer into practice, particularly if e-rubrics and gamification resources were translated into teaching practice by the PSTs.
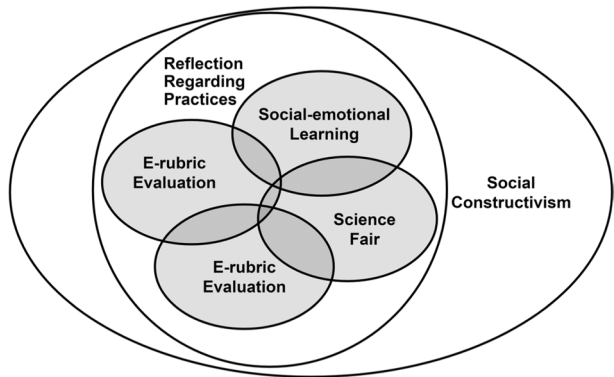
2 Framework

This section describes the key variables considered in this study related to social constructivism, namely reflections regarding practice, gamification resources and science fairs, e-rubric evaluation and social-emotional learning. The combination of all these variables in a social or group setting will favour actual learning and are intended to be a driver for change and the generation of new reflections, ideas and knowledge, which would be insufficient if applied individually (Bovill, 2020). The scheme in Fig. 1 shows these variables.

Key constructs definitions of this research are shown in Table 1.

2.1 Reflections Regarding Practice in Pre-Service Teacher Training

The literature concerning pre-service teacher training emphasises the importance of defining the type of teacher to be trained. From our perspective, pre-service training of secondary school science teachers should contribute to the development of a professional role as a reflexive and critical agent, as defined and discussed in various studies (Danielowich, 2007; Körkkö et al., 2016; O'Keeffe & Paige, 2020; Schön, 1992).

Fig. 1 Proposed reference
framework

Table 1 Key constructs of the research

Key construct	Definition
Social constructivism	A sociological theory of knowledge according to which human develop-ment is socially situated and knowledge is constructed through interaction with others and in which group-based reflection is important (McKinley, 2015; Rannikmäe et al., 2020)
Reflections regarding practice	Reflection should involve self-perception and a reflection of personal competences (including social and personal skills), professional and methodological competences, and in the same state to ensure an integral, holistic reflection of the action-based competence. (Slepcevic-Zach & Stock, 2018)
Gamification resources	Activities involving one or more players, with objectives, rewards, governed by rules and with some degree of competition, and in which social col-laboration is important (Stieglitz et al., 2017)
Science fairs	Activities that make students the protagonist by allowing them to present their scientific project in a context other than the classroom and explain a complex concept to their fellow students (McComas, 2011)
E-rubric evaluation	An evaluation instrument, generally in the form of a grid, although flexible, that collects various evidences and specifies different achievement levels for each (Cebrián-Robles, 2016). E-rubrics have benefits for the teach-ing–learning process such as: a saving in assessment time, more objective feedback and, in short, an aid to student learning (Cebrián-de-la-Serna et al., 2014)
Social-emotional learning	The essential principles of this construct are: (1) Caring relationships are the foundation of all lasting learning, (2) emotions affect how and what we learn, and (3) goal setting and problem solving provide focus, direc-tion, and energy for learning (Elias, 2004)

Reflection is important for both professional teaching and personal development (Le Cornu & Peters, 2005). Thus, Marcelo (2009) considers this initial training period to be very fertile and important as regards learning the teaching profession, and highlights the need to present training proposals to these future teachers aimed at enhancing their reflex-ive, self-critical and self-assessment abilities. These proposals should be formulated con-sidering the concept of the teacher as reflexive practice, with an ability to construct under-standing based on his or her personal and professional involvement.

Since the early work by Schön (1992), the inclusion of references to reflection, reflexive practices and the reflective teacher have become widespread in training programmes and have led to a large number of studies in pre-service teachers in general (Beauchamp, 2015; Van Beveren et al., 2018), and in science education in particular. In the latter case, a reflection in science should be able to convince students of their own explanations for phenomena, which are often counter-intuitive, prior to studying these phenomena (Parker & Heywood, 2013).

Despite this, it is still impossible to conclude that the reflexive teacher is the most common profile in our schools (Russell, 2012). The complexity of changing paradigms, the stability of teaching practices and preconceived ideas arising from observation-based learning are just a few of the obstacles to the development of this type of professional. Moreover, teachers must participate in reflection in order to understand the nature of their teaching, achieve results and see the way in which their personal values and beliefs guide the teaching–learning process, thus helping them to understand the role of education as an instrument for change (Smith et al., 2017). If this atmosphere of reflection can be promoted in PSTs, who then use it in their classrooms once in-service, teaching staff may change from a transition centred on their students to a more in-depth understanding of the content involving an analysis of their scientific beliefs and knowledge (Abell & Lederman, 2007).

In addition, the need to contemplate a broader view of those aspects that must be taken into consideration in PSTs, including the three fields —professional, social and personal— in which various competences must be developed, should be considered (Perrenoud, 2004). As noted by Slepcevic-Zach and Stock (2018), we believe that reflection should involve self-perception and a reflection of personal competences (including social and personal skills), professional and methodological competences, and in the same state to ensure an integral, holistic reflection of the action-based competence.

As part of the personal aspect, we consider it important to reflect on emotions, which can act as facilitators or obstacles to teaching and learning (Blanco et al., 2010; King et al., 2015). Indeed, the study of emotions plays a key role in PSTs (Schoffner, 2009) as they can minimise student tiredness, which is one means of engaging with the most relevant scientific content (Ritchie et al., 2011).

All the above supports the need to integrate reflection into PSTs in the framework of social constructivism, which is the aim of the training programme presented herein, in order to help to overcome various practices which may sometimes appear out of date.

2.2 Gamification Resources and Science Fair, e-rubric Evaluation and Social-Emotional Learning as Elements for Reflection in the Framework of Social Constructivism

2.2.1 Gamification Resources and Science Fair

Although all the elements of the teaching–learning process are susceptible to reflection, in PSTs we consider it important to pay greater attention to the resources to be used and one of the most complex processes to be faced, namely evaluation (Goulette & Swanson, 2019).

Although innovations in educational resources may have great potential for significant learning, they must fulfil the various human needs and material resources available in the classroom, even though these resources often reach schools via commercial networks (McKenney, 2018). In addition to commercial networks, teachers can also use Open

Educational Resources (OERs), which can be found, amongst others, on the internet. The subsequent adaptation of such OERs by teachers to the specific context of their classrooms can improve and redefine the knowledge acquired by them during their studies (Kim, 2018). However, from our perspective, teachers must also be able to develop and innovate their own resources, even though this is not usually standard practice amongst teachers who, in most cases, use science textbooks (Yun & Park, 2018) or simply select materials prepared by other professionals for incorporation into their classes, subsequently adapting them to their own scientific, pedagogical and teaching ideas and beliefs with no in-depth reflection during the process. These aspects have received little attention in educational research and represent an additional difficulty in PSTs (Kang et al., 2016). It is therefore essential to involve teachers in resource design tasks, and their implementation and evaluation, as occurs in research with in-service teachers during the design of teaching sequences to prepare them for innovation (Coenders et al., 2010). This strategy is in agreement with the concept of a teacher as a reflective and innovative professional and also with the social and educational requirements proposed recently.

Amongst the numerous resources available, gamification-based resources (hereinafter GRs), or the use of game design elements in non-game contexts (Deterding et al., 2011), have been widely explored in science teaching (Gaydos & Squire, 2010). GRs are typically activities involving one or more players, with objectives, rewards, governed by rules and with some degree of competition, and in which social collaboration is important (Stieglitz et al., 2017). Some examples of GRs include puzzles (Joag, 2014), board games (Perkins, 2016), card games (Luttikhuizen, 2018), bingo (Tan et al., 2019), augmented reality games (Crandall et al., 2015), video games (Annetta, 2012), etc. The proliferation of GRs is due to the qualities they present in science teaching in the context of a socio-constructivist learning framework. Benefits are seen in terms of the development and promotion of creativity, collaboration, exploration and imagination (Kangas, 2010) and for promoting interpersonal qualities such as respect, fair play, integrity, justice, etc. The implications of the emotional type or the game-playing nature are other aspects that encourage learning. However, perhaps the most important quality is the motivation they produce, which is why the design of this type of approach is valued by teachers as it fits well with their desire to seek new and motivational ways of teaching science. Other researchers argue that the main motivation underlying any GR should be to learn, with this serving as an effective model to do so (Dörner et al., 2016). As such, they may be very effective alternatives for more traditional tasks at all educational levels (Haz et al., 2018).

The format in which the designed GR is presented is also important. Thus, a science fair is a format with significant advantages as it involves teachers in training, thereby improving their performance and facilitating their learning experience (De-Barros-Miller, 2016). Science fairs make students the protagonist by allowing them to present their scientific project in a context other than the classroom and explain a complex concept to their fellow students (McComas, 2011). A science fair provides a common space for the group to allow the generation of shared knowledge (Matusov, 2001). At primary and secondary levels, they provide students with critical thinking skills (De-Barros-Miller, 2016) and promote their creativity and interest in science (Sahin, 2013). The belief that, when used as the main axis of teaching, science fairs are not in alignment with the curriculum is one negative aspect. Despite being widely used in other countries, science fairs have received little attention at a secondary level and in PSTs in Spain, which is why we have decided to use this format to present GRs.

2.2.2 Evaluation Using e-rubrics

Another important aspect that requires reflection from PSTs is evaluation, with one means of improving this capacity involving an "evaluate to learn" approach (Folkes & Carmichael, 2006). In this sense, some proposals concentrate on the interiorisation and reflection of evaluation criteria (Lehesvuori et al., 2017).

Different evaluation methodologies are available to achieve learning by reflection in practice, including self-evaluation activities (Wanner & Palmer, 2018), which allow students to learn strategies that let them self-regulate their learning as well as receive a self-criticism of their work (Steffens & Underwood, 2008). Other such methodologies include peer-evaluation or co-evaluation activities, in which they will carry out a critical exercise with their partner, thus allowing different problem-solving approaches to be appraised (Mustafa, 2017); or the so-called 360-degree evaluation, in which these methodologies are combined (Tee & Ahmed, 2014). This practice-related reflection has been shown to be very useful and effective for improving the performance of pre-service teachers (Lei & Chan, 2018). According to Zimmerman (2013), reflection is the final phase in the cyclic process of self-regulated learning, in which students evaluate and correct their learning before returning to the initial phase of prevision, performance and self-reflection. According to Chetcuti and Cutajar (2014), student learning, which is usually dependent on supervision by the teacher to carry out the evaluations, is required to implement these methodologies.

Although the entire teaching–learning process that takes place in a science classroom is susceptible to evaluation, instruments that allow the degree of compliance with the expected objectives to be assessed objectively are required to carry this out successfully. Of the various evaluation instruments available, we consider rubrics to be of particular interest as they provide benefits in terms of time-saving during evaluation, transmit effective feedback and promote learning (Stevens & Levi, 2005); specifically, electronic rubrics, which facilitate cooperative evaluation (Cebrián-de-la-Serna et al., 2014). In addition, one of the advantages of using online platforms for evaluation is that they promote self-evaluation and autonomous learning that is supervised instantaneously, thereby facilitating dialogue between fellow students, teacher-guided reflection and the communication and justification of the various elements evaluated by the rubric (Cebrián-Robles, 2016). Studies providing the satisfaction results for e-rubrics for teachers and students are available (Quintana et al., 2014).

In light of the above, we have considered reflection in terms of evaluation with e-rubrics in the social-constructivism framework. In addition to being important for PSTs, e-rubrics are one of the tasks they will need to carry out in their future professional career, therefore we are convinced that the collaborative construction of rubrics, and use thereof during evaluation, helps the learning and interiorisation of those aspects that need to be improved to result in a quality instrument. In summary, the involvement of PSTs in evaluation during teaching practices in higher education is relevant as the lecturer is not the only source of evaluation within such processes.

2.2.3 Social-Emotional Learning

For a number of years, researchers in the field of social psychology have studied the influence of mood and emotions on cognitive processing (Fredricks et al., 2016). Emotions are "brief, psychophysiological changes that result from a response to a meaningful situation in one's environment" (Rosenberg, 1998, p. 250) that usually arise in response to a specific

person or event (Linnenbrink & Pintrich, 2002). Emotions are fast, automatic and occur unconsciously yet still have a marked influence on the way in which we think and interpret events (Kagan, 2007). Emotions are also present in science education despite the widely held belief that it is a dispassionate and emotionless discipline, whereas a full range of emotions is actually required for learning (Sinatra et al., 2014). Moreover, collaborative learning promoted by social constructivism may also help to improve the emotions experienced when pursuing common learning. The aim is to engage students in exploring their emotions about each other and about science as well for the purpose of supporting them in improving their social and emotional skills (Matthews, 2004). Different studies in science education provide evidence that positive emotions and enjoyment from learning science play a significant role in learning outcomes and serve as a driving force for self-learning, and for retaining knowledge (Nicolaou et al., 2015). In short, three essential principles of social-emotional learning are: (1) Caring relationships are the foundation of all lasting learning, (2) Emotions affect how and what we learn, and (3) Goal setting and problem-solving provide focus, direction, and energy for learning (Elias, 2004; Haynes et al., 2003).

3 Purpose of the Research

This research was carried out in a training programme for PSTs centred on the design, implementation and evaluation of GRs and e-rubrics for science teaching in the social-constructivism framework and the reflective teacher. The aim was to study the impact of reflection performed by PSTs on the evaluation of GRs when using e-rubrics and optimising them in an iterative process involving all PSTs. The initial hypothesis proposed was that the activities carried out, and the reflection thereof, may influence the criteria proposed when designing the e-rubric or affect their emotions. Secondly, the potential impact of the programme as regards design of the GR and evaluation using e-rubrics may affect its transfer into educational practice.

In order to achieve these goals the following research questions were posed:

- *Research Question 1* What is the impact of reflection on the evaluation of GRs using e-rubrics?: (a) How do the evaluation criteria used by PSTs when designing an e-rubric to evaluate GRs, which will be modified during the programme encouraging reflection, evolve? and (b) how does the group consensus affect the criteria chosen?
- *Research Question 2* What is the impact of emotions on PSTs when participating in reflection activities related to the design of GRs and an e-rubric to evaluate them?
- *Research Question 3* What is the impact of the programme on transfer into practice?: (a) How do PSTs perceive the e-rubric as an evaluation instrument before and after participating in the programme? and (b) to what extent are the PSTs able to transfer the knowledge that they have learned in this training programme into practice?

4 Method

4.1 Participants

A nonprobability convenience of 50 Spanish PSTs studying the "Teaching innovation and introduction to educational research" module of the Masters in Secondary School Teaching

at the University of Málaga (Málaga, Spain) in the academic year 2018/19 was taken. This subject requires a reflection concerning educational innovations as one of the main teaching contents as they are being trained for their professional activity as teachers. These PSTs were from the specialist subjects Biology and Geology (31) and Physics and Chemistry (19). Of these, 60% were female and the rest male, and they were aged between 21 and 39 years. The most common profile amongst Biology and Geology PSTs was a degree in biology or environmental science, whereas that for Physics and Chemistry was a degree in chemistry or chemical engineering. It should be noted that although participation in all activities was high, as many as nine participants failed to answer some questionnaires. Moreover, 53.1% of PSTs had never used a rubric for evaluation, and 97.9% had never used an e-rubric.

The programme was designed jointly by the paper's three authors, who had extensive experience as lecturers in science education. Two lecturers (second and third authors) collaborated with the course lecturer (first author) in the implementation and the evaluation.

4.2 Training Programme

This programme was carried out during this module as part of a topic on educational innovation in science teaching that included a series of tasks which allowed PSTs to reflect on GRs and e-rubrics. The programme comprised seven sessions (8.5 h of class participation plus 25 h of personal homework: 20 h for design and construction of the GR and 5 h for design and improvement of the e-rubrics) distributed over a period of 23 days.

Figure 2 provides a schematic view of the training programme, showing the four dimensions that were carried out in parallel: (1) Reflection activities; (2) GRs, in which PSTs received training and designed, constructed and exchanged experiences; (3) evaluation
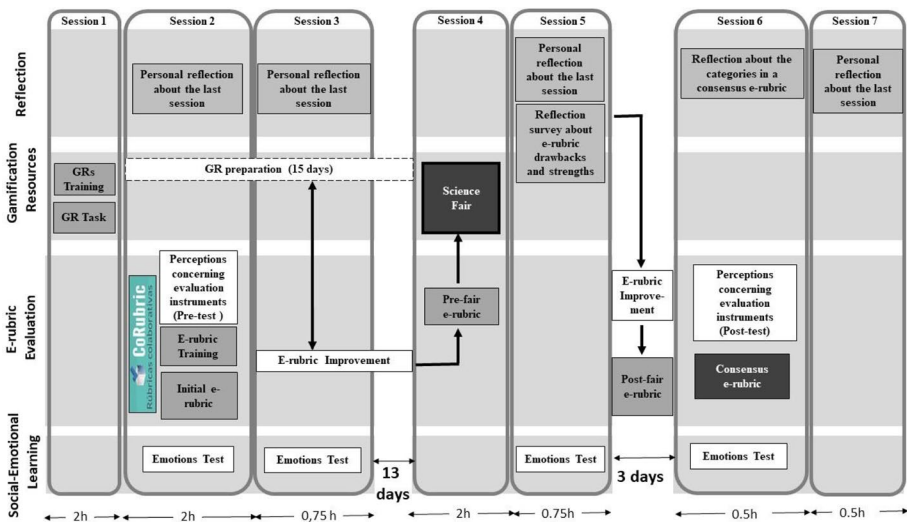

Fig. 2 Training programme

using an e-rubric, in which PSTs were expected to design, optimise and use their own e-rubric to evaluate GRs, and then optimise it with the group; (4) social-emotional learning, in which students were expected to question their own emotional situation in each of the activities in order to learn how to understand their emotions as a future teacher.

4.2.1 Dimension 1: Reflection Activities

This dimension was mainly carried out via a training activity known as personal reflection (Luque et al., 2021), which was obligatory and held at the start of the session. It is considered to be the presentation of a personal view of a topic covered during the module and involves summarising, self-evaluation, broadening of understanding and the drafting of questions. All PSTs were required to perform this once during the module. At each session, two PSTs reflected independently about aspects covered in the previous session, using a digital presentation of around 10 min and including three important ideas covered in class, three aspects that they did not understand, three aspects they would have liked to learn more about, and one question for the discussion. As such, PSTs reflected on training in GRs and e-rubrics in sessions 2 and 3, respectively, on the GR fair and the efficacy of the e-rubric designed in session 5, and on the consensualised e-rubric in session 7. In addition, two surveys were used in sessions 5 and 6 to make them reflect on some aspects of the e-rubric, namely its drawbacks or strengths, or consensus criteria.

4.2.2 Dimension 2: Gamification-Based Resources

This dimension involves training, design and elaboration, implementation and dissemination of a science fair as part of a GR.

Gamification-Based Resources Training (session 1) The aim was to determine best practices for GRs for use in secondary education and analyse key aspects of the design, implementation and evaluation thereof. The lecturer presented different GRs, supported by research, previously implemented in a secondary school classroom. During the session, PSTs had the opportunity to use and analyse, amongst others, a set of cards for learning the periodic table (Franco-Mariscal et al., 2012), a puzzle to discover the bones in the body (Franco-Mariscal & Cano-Iglesias, 2011), augmented and virtual reality tools for science (Moreno & Franco-Mariscal, 2019), a worldwide football competition based on the properties of the chemical elements (Franco-Mariscal, 2014) or a bingo concerning the properties of atoms (Franco-Mariscal & Cano-Iglesias, 2014) (Fig. 3).

Design and Elaboration of the Gamification-Based Resources The design and elaboration of a GR for secondary education (12 to 18 years) were proposed as a task (session 1), with two products having to be submitted: The GR itself and a project explaining the design, including learning objectives, the educational level at which it is to be used, the competences to be developed, the scientific aspects covered, the materials required, methodology, instructions for use and explanatory photographs. The PSTs designed board games (13), cards (10), puzzles (8), digital GRs (7), question and answer games (3), dominoes (2), a role-playing game (1) and an escape room (1). Some examples can be found in Fig. 4. Despite having asked for a GR, five PSTs designed resources that did not meet this requirement.

Implementation of the Gamification-Based Resources in a Science Fair The PSTs presented their GR in a science fair in class (session 4), 15 days after session 1, and were given access to a stand to present the GR and a poster; they were allowed to use any type of
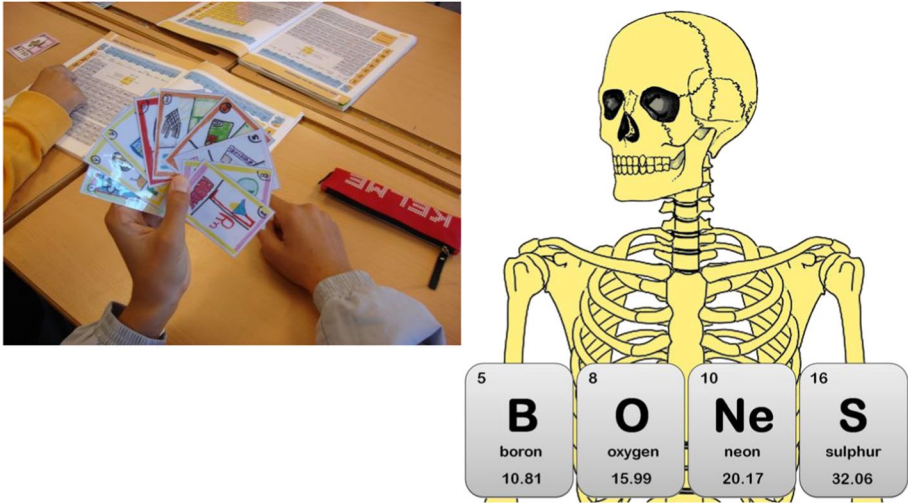
Fig. 3 Some of the GRs used in the training session: set of cards for the periodic table (left) and skeleton puzzle (right)

Fig. 4 Examples of GRs designed by the PSTs: cards to learn the periodic table (top left), board game on blood groups (top right) and digital resource on the cardiovascular system (bottom)

material. The fair was held in two sittings of one hour each, with the PSTs being organised into two groups with different roles. At each sitting, one group of PSTs played the role of the teacher implementing the GR, explaining it to their fellow students (second group), who played the role of a student visiting the fair and evaluating the GRs (Fig. 5). These groups exchanged roles after the first hour.

4.2.3 Dimension 3: Evaluation Using e-rubrics

The third dimension of the programme involved the e-rubric as an evaluation instrument and included training, different design activities and optimisation of the e-rubric for evaluating GRs. It comprised the following activities:

(1) Perceptions of Evaluation Instruments (sessions 2 and 6). The PSTs responded to the question "With which evaluation instrument(s) will you feel most comfortable evaluating your future students: a written exam, observation in class, portfolios, rubrics, questionnaire, oral exam, essays or evaluation with GRs? You can choose more than one instrument".

(2) Training in e-rubrics and CoRubric (session 2) This training covered e-rubrics as an evaluation instrument and the free collaborative evaluation platform (corubric.com) (Cebrián-Robles, 2016). It was noted that an e-rubric is an evaluation instrument, generally in the form of a grid, although flexible, that collects various evidences and specifies different achievement levels for each (Cebrián-Robles, 2016). CoRubric was presented as a collaboration-based methodology and technological tool that is characterised by its flexibility, and allows e-rubrics with evidences that present different levels of achievement to be designed. This tool allows different qualitative and quantitative evaluation methodologies to be designed and applied, such as peer evaluation, self-evaluation, ipsative evaluation, 360° evaluation or group evaluation. It allows evaluation using mobile devices, and the possibility of receiving other evaluations performed by fellow students to be received in situ. As part of this training, instructions regarding how to register with the CoRubric platform and how to create e-rubrics were provided, along with the means of including it in a common project that allows each PST to see the e-rubrics of their fellow students and use their own e-rubric to evaluate GRs. PSTs had the opportunity to use an e-rubric to evaluate the personal reflection activity performed by a fellow student.


Fig. 5 Science fair for GRs in the classroom

(3) Design of an Initial e-rubric to Evaluate Gamification-Based Resources This task consisted of preparing an *initial e-rubric* using CoRubric, which each PST then had to improve, on an individual basis, on at least three occasions during the programme. This was posed as follows: "Using CoRubric, construct an e-rubric that can be used to evaluate all GRs that will be created by your fellow students. This e-rubric must contain a minimum of three elements and a maximum of five. Each element must have four achievement levels." When constructing the *initial e-rubric*, the PSTs had not yet prepared the GR, therefore the design concepts corresponded to their misconceptions.

(4) Construction of a Second e-rubric (Pre-Fair e-rubric) Based on Reflection During the Preparation of a Gamification-Based Resource Whilst the PSTs were preparing the GR, they were given a perfection time of 14 days in which they could modify the *initial e-rubric*, if they considered this to be necessary, to include aspects related to the GR. During this period, PSTs did not receive any instructions concerning how to improve the *initial e-rubric*, although they were given access to the e-rubrics designed by their fellow students via CoRubric and could search for information on the internet. Moreover, session 3, which involved a reflection on e-rubric training, was also held in this period.

(5) Use of the Pre-Fair e-rubric to Evaluate Gamification-Based Resources in a Science Fair The PSTs were given the opportunity to use CoRubric and the *pre-fair e-rubric* to evaluate the GRs of other PSTs at the science fair (session 4; Fig. 6).

(6) Construction of a Third e-rubric (Post-Fair e-rubric) After the fair, PSTs were able to consult the evaluations given to their GRs by their fellow students in CoRubric. This activity allowed them to reflect on the drawbacks and strengths of their e-rubric in situ, doing so in writing by completing a survey (Table 2) (session 5). The PSTs were given three days to prepare this *post-fair e-rubric* as a result of this reflection.

Fig. 6 PST using CoRubric to evaluate a GR at the fair
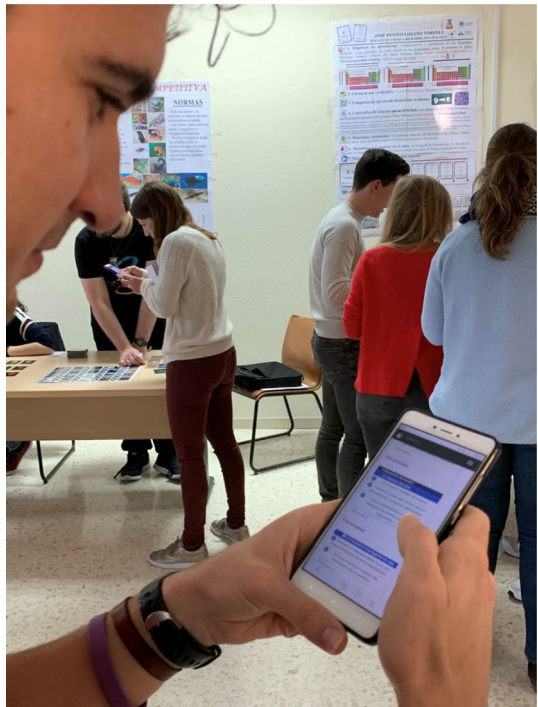
Table 2 Survey to reflect on the efficacy of the *pre-fair e-rubric*

1. In the fair you have used the e-rubric you created in CoRubric to evaluate the GRs of your fellow students. Do you think it needs to be improved?

2. (a) I think that my e-rubric is: ☐ objective ☐ subjective. (b) State why you chose that answer

3. What do you think are the strengths and drawbacks of your e-rubric when evaluating the GRs at the fair? Please explain

(7) Construction of a Consensus e-rubric At session 6, and in light of the results of the *post-fair e-rubric*, the PSTs performed the following task, which allowed the *consensus e-rubric* to be constructed as a group: *In this activity, we are all going to decide on a consensus e-rubric to evaluate a GR. To that end, the course lecturers have reviewed the e-rubrics designed by you. Amongst all PSTs, you have used a total of 18 items which you consider to be important for evaluating a GR. Rate each one from 1 to 10 depending on their suitability for inclusion in a consensus e-rubric for evaluating a GR.*

4.2.4 Dimension 4: Social-Emotional Learning

Upon completing each session, PSTs were each asked to complete an adapted version of the KPSI (Knowledge and Prior Study Inventory) questionnaire (Jiménez-Liso et al., 2021) (Appendix), in an attempt to get them to reflect on the emotions experienced in some activities of the programme, indicating which emotions they had felt among the nine presented, and were able to evaluate as many as they wished. Five important aspects concerning GRs (training and fair) and the e-rubrics (training, use and consensus) were covered. This questionnaire uses different emotions applicable to educational activities while avoiding an overlap between them, specifically: rejection, concentration, insecurity, interest, boredom, confidence, satisfaction, dissatisfaction, and shame. This study is not intended to stimulate some emotions or consider some as positive (e.g. interest, satisfaction) and others as negative (e.g. insecurity, boredom), a dichotomy that has been used to analyse teaching practice (Marks, 2000) by associating positive emotions with success and negative emotions with failure (Pekrun, & Linnenbrink-García, 2014). In our opinion, emotions are not always positive or negative but are much more complex and are found on a continuum. This study aims simply to recognise the emotions experienced by way of an individual and collective reflection designed to describe which types or emotions are experienced in the classroom setting during the programme while giving a meaning to them as part of educational practice.

The questionnaire also included a question to evaluate their perception of understanding, before and after the session, of the aspects covered, using a Likert scale with responses from 1 to 5 points (1: I do not know anything, 2: I know a little, 3: I know it well, 4: I know it very well, 5: I can explain it to a friend).

4.3 School-Based Practicals for Pre-Service Secondary Science Teachers

The participants in this study carried out 125 h of practice in a secondary school over a period of five months, which started one month after the end of the proposed training programme. During this period, the PSTs had to design and implement a science teaching–learning sequence for secondary school students. At the end of the period, they

presented a portfolio in which this sequence was included. These teaching–learning sequences were not predefined in the training programmes of their Masters course and could include content from any subject. Thus, the PSTs themselves decided the specific work topic, the content and their approach to teaching activities, although the content was always agreed with their academic and professional tutors, who were not involved in the research presented here. Consequently, PSTs were free to decide whether or not to include GRs designed by themselves or e-rubrics to evaluate as part of the teaching–learning sequence. In addition to these practicals, PSTs were required to submit a Final Masters Project, which included the teaching–learning sequence designed and implemented and an improved version of that project based on the reflection of their teaching practice. The PSTs submitted the Final Masters Project six months after completing the training programme. For this study, it was only possible to access the 38 Final Masters Projects deposited in the official repository established by the faculty.

5 Data Analysis

Data analysis was carried out in three stages, one for each research question proposed. The first study analysed the impact of reflection on the evaluation of the GRs by studying the evolution of the four e-rubrics designed by the PSTs during the programme (Research Question 1). The second analysis involved emotions and the evolution of understanding at different moments associated with the GR and the e-rubric (Research Question 2). The third analysis addressed transfer into practice, concentrating on the design of the GR and use of e-rubrics found in the PSTs' Final Masters Project (Research Question 3).
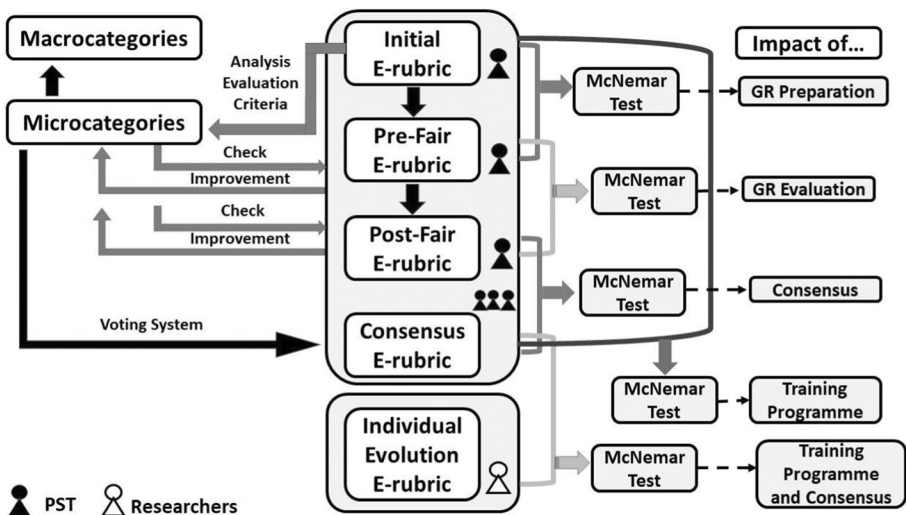

Fig. 7 Data analysis of e-rubrics

5.1 Data Analysis 1: E-rubrics

The researchers established the items that the PSTs considered important for assessing the GRs by performing the analysis shown in Fig. 7.

5.1.1 Establishment of Micro- and Macrocategories

Initially, the evaluation criteria proposed by each PST in the *initial e-rubric* were analysed, establishing microcategories that included statements regarding the same idea. It was also determined whether these microcategories were maintained in *the pre-fair e-rubric* and *the post-fair e-rubric*, or whether new ones appeared after the reflection activities carried out. Finally, these microcategories were grouped into two macrocategories to differentiate between who (speaker) or what (GR) was the basis for each evaluation criterion.

5.1.2 Establishment of a Consensus E-Rubric

The microcategories were used to allow the PSTs to establish to what extent each evaluation criterion was pertinent for each evaluation criterion in the collective *consensus e-rubric*. A score-based voting system was used, and each PSTs could score each microcategory from 1 to 10 points (Section 4.2.3, point 7). The total score given by the 50 PSTs to each microcategory was then added up, with the score ranging from 50 to 500 points (if all PSTs had scored that microcategory as 1 or 10, respectively). The researchers selected half of the microcategories (those with the highest scores) for the *consensus rubric*.

5.1.3 Evolution of the Categories in the E-Rubrics

To study the evolution of the evaluation criteria from the *initial e-rubric* to the *consensus e-rubric*, the frequency of appearance of each microcategory at the four stages was compared. Specifically, the number of PSTs who included each microcategory was analysed for the different possible combinations of the four e-rubrics. These possible combinations were represented with a four-digit code consisting of 0 and 1 (for example 1010), with each position representing the corresponding e-rubric (the first digit relates to the *initial e-rubric*, the second to the *pre-fair e-rubric*, and so on), with the value of 0 or 1 indicating whether said category appeared in that e-rubric or not, respectively. Thus, the combination 1010 assigned to a specific microcategory indicates that said criterion was only included in the *initial e-rubric* (1 in the first digit) and the *post-fair e-rubric* (1 in the third digit).

The McNemar test was used to check for possible statistically significant differences between the different moments at which PSTs prepared the e-rubric as these are categoric samples of related non-parametric variables, in other words data that do not fit a normal curve. Specifically, three key moments in the programme that could affect the decision taken by the person preparing the e-rubric, and which allowed three pairs of e-rubrics to be compared, were chosen, namely:

- Preparation of the GR: This allows a comparison between the *initial e-rubric* and the *pre-fair e-rubric*.
- Use of the e-rubric to evaluate GRs at the science fair: This allows a comparison between the *pre-fair e-rubric* and the *post-fair e-rubric*.

- Establishment of the consensus e-rubric: This allows a comparison between the *post-fair e-rubric* and the *consensus e-rubric*.

In addition, the McNemar test was also used to check the impact of the training programme. This test allows a comparison between the *initial e-rubric* and the *consensus e-rubric*.

Finally, the researchers in this study constructed a new e-rubric (*individual evolution e-rubric*) to study the evolution between the first three e-rubrics reflected on individually by the PSTs and the *consensus e-rubric* reflected on collectively. This *individual evolution e-rubric* considers all the microcategories which the students highlighted in at least one or their three individual rubrics. The McNemar test was also used to check for statistically significant differences between the *individual evolution e-rubric* and *the consensus e-rubric*. The frequency of PSTs who used each microcategory in at least one of the three e-rubrics prepared was taken into consideration for the *individual evolution e-rubric*.

5.2 Data Analysis 2: Emotions and Evolution of the Perception of Understanding.

Emotions were analysed quantitatively based on the PSTs' frequency who expressed each emotion in the most important aspects concerning GRs (training and fair) and the e-rubrics (training, implementation and consensus). It was taken into account that the PSTs could mark more than one emotion. Therefore, a bar graph was constructed representing each bar as an activity of the training programme.

The perception of the understanding of GRs and e-rubrics before and after each activity was analysed, calculating the percentage of PSTs who marked each value on the Likert scale (from 1 to 5) in the two moments. Finally, a comparison between the majority percentages in each case was performed to detect any progress that had taken place.

5.3 Data Analysis 3: Transfer into Practice.

Analysis of the possible change in perception of the PSTs regarding the use of the e-rubric as an evaluation instrument as a result of the programme was performed quantitatively through the variation in the percentage of their choice before and after the training programme.

Data analysis for the transfer into practice in the Final Masters Projects was performed as follows. Initially, a search was carried out for possible references to the design of GRs and use of e-rubrics in the Final Masters Project by reading each portfolio. In addition, portfolios were submitted in electronic format (PDF), and a search was carried out using Adobe Acrobat Reader to confirm that all the references to GRs and e-rubrics had been analysed. In this search, the keywords "resource", "fair" and ''rubric'' were used to allow the different words related to GRs and e-rubrics to be located. The second step involved an analysis of the extent to which the GRs or e-rubric were mentioned in the Final Masters Project. This was achieved by establishing a system of progress that had different categories for the ideas learned during the training programme. For the case of GRs: (1) GR is not mentioned in the Final Masters Project; (2) GR is mentioned, but is not part of the teaching–learning sequence activities; and (3) GR is mentioned and is part of the design of activities for the teaching–learning sequence and is put into practice in the intervention. For the e-rubric: (1) rubric is not mentioned in the Final Masters Project; (2) rubric is mentioned, but is not part of the teaching–learning sequence design; (3) rubric is mentioned

and is incorporated into the teaching–learning sequence design but is not involved in the intervention; (4) rubric is mentioned, is incorporated into the teaching–learning sequence design and is involved in the intervention, but it is not an e-rubric; and (5) rubric is mentioned, is incorporated into the teaching–learning sequence design, is involved in the intervention and an e-rubric is used.

6 Results and Discussion

6.1 Results 1: E-Rubrics

6.1.1 Establishment of Micro- and Macrocategories

Table 3 shows the micro- and macrocategories categorised by the authors of this study in accordance with the evaluation criteria used by the PSTs in their e-rubrics.

The first 15 microcategories (Table 3) were identified during analysis of the *initial e-rubric*. Three new microcategories (M16: Time required to use the GR; M17: Ability of the GR to transmit the content; and M18: Ability of the GR to evaluate understanding) emerged in the *pre-fair e-rubric*. The improvements made to the *post-fair e-rubric* and the *consensus e-rubric* did not generate new microcategories, which remained at 18. All microcategories could be grouped into two macrocategories, one centred on the speaker for the GR at the fair and the other centred on the GR itself.

Table 3 Micro- and macrocategories

Microcategories	Macrocategories
M01. Oral explanation of the GR	Speaker
M02. Interaction with the public	
M03. Understanding of the topic	
M04. Adaptation of the GR to the curriculum	GR
M05. Adaptation of the GR to the educational level	
M06. Utility of the GR for learning	
M07. Ability of the GR to motivate	
M08. Sources used to prepare the GR	
M09. Quality of GR preparation	
M10. Ability of the GR to develop scientific competences	
M11. Relationship between GR and teaching in context	
M12. Creativity/originality/innovation of the GR	
M13. Ease of use of the GR	
M14. Interactiveness of the GR	
M15. Versatility of the GR for use in other activities	
M16. Time required to use the GR	
M17. Ability of the GR to transmit the content	
M18. Ability of the GR to evaluate understanding	

GR Gamification-based resources

Table 4 Evolution of the e-rubric for one PST

Initial e-rubric:

Content of the GR			
1 The GR is poorly described and does not have a clear objective	2 The description is sufficiently well documented but does not comply with the learning objective	3 The description is well documented but barely mentions the learning objective	4 The description is very complete and documented. It also complies with the learning objectives

Pre-fair e-rubric:

Quality of GR content			
1 The topic covered by the GR is not appropriate, is not well documented or described, or does not provide information about it	2 The topic of the GR is acceptable but there is little documentation, or the documentation provided is inadequate	3 The GR is well documented and described and concerns a relevant topic. This topic is covered clearly and logically	4 The GR is very well documented and described in a very detailed manner. The topic covered is appropriate, suitable and relevant

Post-fair e-rubric:

3.Quality of GR content			
1 As for pre-fair e-rubric + The methodology is not appropiate	2 As for pre-fair e-rubric + The methodology could be appropiate	3 As for pre-fair e-rubric + The methodology is appropiate	4 As for pre-fair e-rubric + The methodology is very appropiate

As an example, Table 4 presents the evolution of the evaluation criterion *content of the GR*, associated with M08 (Sources used to prepare the GR) and proposed by one of the PSTs at the three perfection stages.

As can be seen, this criterion was modified in terms of both the text defining it and the different achievement levels, and the process of improving the e-rubric provided aspects to be taken into consideration in these levels.

In general, the most important changes were found between the *initial e-rubric* and the *pre-fair e-rubric* and concerned improvements to the wording of the indicators and achievement levels for the various criteria similar to those of the examples. These changes were performed as a result of the reflection process carried out by the PSTs and may have been influenced by the preparation of their own GR, access to the *initial e-rubric* of their fellow students in CoRubric, the search for information on the internet or the personal reflections of PSTs in class. In contrast to Chetcuti and Cutajar (2014), who considered supervision of rubric preparation by the tutor to be essential, in the design of e-rubrics, PSTs did not require such supervision when modifying the evaluation instrument based on the limitations and reflections they performed during the training programme.

Similarly, the improvements made between the *pre-fair e-rubric* and the *post-fair e-rubric* were much less significant and, indeed, comprised the incorporation of minor variations in the criteria and indicators formulated previously rather than the introduction of new microcategories. The analysis of the reflections carried out by the PSTs at that stage, by way of a strengths and drawbacks survey for the *pre-fair e-rubric* after use in the science fair (Table 2), supports these minor changes. Thus, 98% of PSTs indicated that, after

being used at the fair to evaluate the GRs of their fellow students, the *pre-fair e-rubric* needed to be improved, and 67.3% considered their e-rubric to be subjective. However, the majority were unable to apply this evaluation in the *post-fair e-rubric* and justified the need for changes in one of the following ways: (1) the e-rubric is too long for the time anticipated to evaluate a GR (6% of PSTs); (2) the achievement levels are imprecise (31%); (3) the e-rubric requires a higher number of evaluation criteria (41%); and (4) I disagree with the score that other e-rubrics give my GR (4%). As can be seen, the reflections noted by the PSTs simply reflect weaknesses in the *pre-fair e-rubric* for some criteria rather than a need to change the type of microcategory. In addition, it can be seen that all these changes are related to the macrocategory GR, with no changes being proposed for the macrocategory speaker.

6.1.2 Establishment of a Consensus E-Rubric

Table 5 lists the total score for each microcategory and indicates those accepted for inclusion thereof in the *consensus e-rubric*. Those microcategories that scored equal to or above 346 points, which was the score of the ninth microcategory in the ranking, were selected.

These results are used in the following section to perform a statistical treatment that will allow us to compare the evolution of the evaluation criteria selected by each PST and the consensus version, thereby comparing individual and collective thinking.

6.1.3 Evolution of the Categories in the E-Rubrics

Table 6 lists the PSTs' appearance frequency (f) (from highest to lowest) of the 18 microcategories in the possible combinations of the four e-rubrics. For instance, for the combination 0110, 12 PSTs included the microcategory M08 in the intermediate e-rubrics (*pre-fair e-rubric* and *post-fair e-rubric*) only. The total for each possible combination for the appearance of microcategories in e-rubrics gives some idea of the impact of the different activities in the programme as regards changes to the evaluation criteria as a result of reflection. The combinations with a high frequency in a microcategory, corresponding to $f \geq 22$ (50% or more of the participating PSTs) are highlighted in black, those with $11 < f < 22$ (between 25 and 50% of PSTs) in grey and those with $f < 11$ (less than 25% of PSTs) in white.

As can be seen (Table 6), the combination 0000 is the most common and comprises four microcategories not assigned to any of the e-rubrics by more than 50% of the PSTs, including the *consensus e-rubric*: *Sources used to prepare the GR* (M08), *Relationship between the GR and teaching in context* (M11), *Versatility of the GR for use in other activities* (M15), and *Time required to use the GR* (M16). Indeed, M16 is the least common microcategory, being absent from the e-rubrics of more than 60% of PSTs.

It should also be noted that the second most common combination (0001) represents the appearance of some microcategories in the consensus version only. The categories *Adaptation of the GR to the educational level* (M05) and *Ability of the GR to evaluate students' understanding* (M18) had the greatest impact on the changes in the *consensus e-rubric* given that more than 50% of PSTs had not even taken these criteria into consideration during the e-rubric construction process (from the *initial e-rubric* to the *post-fair e-rubric*).

Of those microcategories taken into consideration in all e-rubrics (combination 1111), the *Creativity/originality/innovation of the GR* (M12) was the only one considered by more

Table 5 Microcategories with or without consensus for the *consensus e-rubric*

Microcategory	Total Score	Consensus*	Microcategory	Total Score	Consensus*
M01. Oral explanation of the GR	351	Yes	M10. Ability of the GR to develop scientific competence	343	No
M02. Interaction with the public	327	No	M11. Relationship between GR and teaching in context	299	No
M03. Understanding of the topic	347	Yes	M12. Creativity/originality/innovation of the GR	369	Yes
M04. Adaptation of the GR to the curriculum	379	Yes	M13. Ease of use of the GR	374	Yes
M05. Adaptation of the GR to the educational level	382	Yes	M14. Interactiveness of the GR	346	Yes
M06. Utility of the GR for learning	403	Yes	M15. Versatility of the GR for use in other activities	312	No
M07. Ability of the GR to motivate	399	Yes	M16. Time required to use the GR	265	No
M08. Sources used to prepare the GR	188	No	M17. Ability of the GR to transmit the content	324	No
M09. Quality of GR preparation	314	No	M18. Ability of the GR to evaluate understanding	337	No

*Consensus is considered with half of the microcategories having higher scores (≥ 346)

Table 6 Frequency of PSTs in each microcategory for the possible combinations of e-rubrics

| | PSTs' frequency (f) in each microcategory | | | | | | | | | | | | | | | | | | |
E-rubrics combinations*	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	M13	M14	M15	M16	M17	M18	Total
0000	11	15	12	3	2	0	2	27	11	10	28	5	5	11	22	31	15	19	229
0001	9	20	17	18	22	10	12	1	4	13	10	0	11	15	12	3	20	23	220
0111	9	3	6	11	12	10	6	1	4	10	2	6	10	3	3	5	5	1	107
1111	10	0	2	2	4	20	14	0	5	0	1	22	11	1	2	0	0	0	94
0110	1	1	3	3	1	0	1	12	6	6	0	2	4	4	3	4	3	0	54
1001	2	2	3	4	1	2	7	0	5	1	0	5	2	8	0	0	0	0	42
1000	0	1	0	1	0	1	1	1	5	3	1	2	0	1	1	0	0	0	18
1110	1	1	0	0	0	0	0	1	3	0	0	1	0	0	0	0	0	0	7
0011	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0010	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1010	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1

* Each position is representing the corresponding e-rubric (first digit: *initial e-rubric*, second digit: *pre-fair e-rubric*; third digit: *post-fair e-rubric*; fourth digit: *consensus e-rubric*), with the value of 0 or 1 indicating whether said category appeared in that e-rubric or not, respectively.

than 50% of PSTs. The remaining combinations are not representative as they were not considered by more than 50% of PSTs.

Table 7 presents the frequency of appearance of each microcategory in each e-rubric and in at least one of the first three (*individual evolution e-rubric*). For example, 21 PSTs included M01 in at least one of the *initial e-rubric*, the *pre-fair e-rubric* or the *post-fair e-rubric*. The results of the McNemar test carried out to detect possible statistically significant differences between the e-rubrics designed at different moments of the programme: preparation of the GR (*initial e-rubric* & *pre-fair e-rubric*), evaluation of the GR in fair (*pre-fair e-rubric* & *post-fair e-rubric*), consensus (*post-fair e-rubric* & *consensus e-rubric*), training programme (*initial e-rubric* & *consensus e-rubric*) and during the programme and after consensus (*individual evolution e-rubric* & *consensus e-rubric*). The last column in Table 7 lists the differences between the evolution of the e-rubrics for each *individual evolution e-rubric* and the *consensus e-rubric*.

These results showed that some aspects related to the design and preparation of the GR had a marked influence on the selection of criteria between the *initial e-rubric* and the *pre-fair e-rubric* given that statistically significant differences were found for three microcategories (M05: *Adaptation of the GR to the educational level* [$\chi 2 = 8.3$; $p = 0.006$], M08: *Sources used to prepare the GR* [$\chi 2 = 6.8$; $p = 0.012$] and M13: *Ease of use of the GR* [$\chi 2 = 7.1$; $p = 0.013$]). This may be due to the fact that, during preparation of the GR, PSTs had to reflect individually on the design of their own GR and the best way of evaluating learning with other GRs of which they were unaware prepared by their fellow PSTs. The fact that they were expected to construct their GR individually showed them that some important characteristics (educational level, sources or ease of use) favoured learning and could be extrapolated to other GRs. This reflection fits well with the constructivist approach, as the construction of knowledge begins with students, who actively participate in the construction of their own learning (Poerksen, 2004; Woolfolk, 2014). The case of M08 caused a high percentage of PSTs to reconsider its inclusion in the e-rubrics at the stage at which PSTs were collecting information for preparation of their own GR although,

Table 7 Frequencies for microcategories in each e-rubric and McNemar test applied to different stages in the programme

E-rubrics comparison moments

Micro-category	Frequency					BA elaborating the GR		BA assessing the GR		BA consensus		BA training programme		During the programme and after consensus	
	R1	R2	R3	R4	R5	R1 & R2		R2 & R3		R3 & R4		R1 & R4		R5 & R4	
						χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
M01	13	20	20	24	21	NS		NS		NS		9.3	0.003	NS	
M02	4	5	5	18	7	NS		NS		9.9	0.002	10.9	0.001	5.8	0.038
M03	5	9	9	23	11	NS		NS		9.8	0.003	18.0	0.000	NS	
M04	7	14	15	29	20	NS		NS		9.8	0.003	20.2	0.000	20.2	0.000
M05	5	15	16	33	17	8.3	0.006	NS		15.2	0.000	28.0	0.000	16.0	0.000
M06	22	28	28	35	29	NS		NS		7.0	0.016	11.3	0.001	6.0	0.031
M07	22	18	18	32	23	NS		NS		12.3	0.001	8.3	0.006	5.4	0.035
M08	2	11	11	1	12	6.8	0.012	NS		8.3	0.006	NS		9.3	0.003
M09	17	16	16	18	23	NS		NS		NS		NS		NS	
M10	4	11	11	19	18	NS		NS		NS		10.7	0.001	NS	
M11	3	2	3	10	5	NS		NS		5.4	0.039	NS		NS	
M12	29	26	26	29	33	NS		NS		NS		NS		NS	
M13	12	22	22	32	22	7.1	0.013	NS		7.1	0.013	20.0	0.000	8.3	0.006
M14	9	6	6	22	16	NS		NS		11.6	0.001	11.3	0.001	NS	
M15	3	6	6	15	8	NS		NS		6.2	0.022	10.3	0.002	NS	
M16	0	7	7	8	9	–		NS		NS		–		NS	
M17	0	6	6	20	6	–		NS		NS		–		10.9	0.001
M18	0	1	1	21	1	–		NS		NS		–		20.0	0.000

R1: *initial e-rubric*; R2: *pre-fair e-rubric*; R3: *post-fair e-rubric*; R4: *consensus e-rubric*; R5: *individual evolution e-rubric*

NS no significative; p: value signification with $p < 0.05$

BA before and after

as noted above, more than 50% of them decided not to include it in any of the four e-rubrics in the end (combination 0000).

Evaluation of the GRs at the fair using an e-rubric did not affect the changes produced between the *pre-fair e-rubric* and the *post-fair e-rubric*, thus demonstrating that reflection of the evaluation criteria during preparation of the GR was sufficient. Despite this, use of the *pre-fair e-rubric* demonstrated the need to redefine the tests and achievement levels in the criteria, as also shown by the reflection survey completed after use of the e-rubric.

However, the programme stage with the most marked influence was clearly the contribution of criteria to the group when establishing the consensus version, as statistically significant differences were detected for 11 microcategories between both the *pre-fair e-rubric* and the *consensus e-rubric,* and between the *initial e-rubric* and the *consensus e-rubric*. These findings appear to indicate that a global vision of the criteria proposed by the group of PSTs causes them to reflect and to make new changes to their own e-rubrics.

Finally, the comparison between the *individual evolution e-rubric* and the *consensus e-rubric*, which showed statistically significant differences for 50% of the microcategories, should be noted. This highlights the fact that the criteria in an e-rubric used to evaluate GRs are different if designed in a consensual manner in a group or individually. In other words, PSTs have greater confidence when selecting criteria from a list generated in consensus with their fellow students than those that they would have chosen previously, which they exclude from the consensus despite their presence in the three previous e-rubrics. These findings reinforce the idea that encouraging PST to discuss ideas and share them with their fellow students leads to better learning regarding complex tasks. In this case, the consensus criteria for evaluating an e-rubric is a clear example of social constructivism (Akpan & Kennedy, 2020).

6.2 Results 2: Emotions and Evolution of the Perception of Understanding

Figure 8 shows both the emotions felt by the participants and the frequencies of each. The first two columns relate to GR-related aspects and the final three to the e-rubric.

In general, it can be seen that not all emotions were experienced equally by the PSTs. Thus, several emotions are more common than others given the high frequencies obtained. This is the case, for example, for interest, satisfaction and concentration, with the former being the most common emotion at all stages analysed. Training in GRs was well-received,

Fig. 8 Emotions experienced by the participants
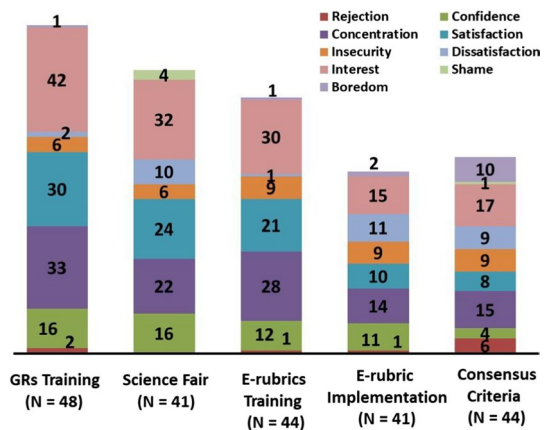
with a high degree of interest (42/48 PSTs) and concentration (33/48). Once the GR had been designed and implemented, this interest was maintained (32/41) and also generated satisfaction (24/41). In the case of e-rubrics, training also generated interest (30/44) and concentration (28/44), although their use at the fair generated low levels of interest (15/41), satisfaction (10/41) and confidence (11/41), thus suggesting that, at that stage, it was more important for PSTs to discover new GRs than to evaluate them. The implementation of the e-rubric also resulted in some degree of insecurity (9/41), thus highlighting that many PSTs did not agree with the result of their e-rubric. Finally, although the consensus criteria produced some degree of interest (17/44) and concentration (15/44), they also generated boredom (10/44) and insecurity and dissatisfaction (9/44), thus appearing to indicate that some PSTs accepted the group consensus but continued to believe that some of the criteria not present in the *consensus e-rubric* should have been included.

Table 8 lists the perception of understanding (in percentages) that PSTs indicated before and after these five activities. In all cases, an evolution of their learning of both GRs and e-rubrics was observed as the majority responses were found prior to the activities for options 1 and 2 and after them for options 4 or 5.

6.3 Results 3: Transfer into Practice

The programme resulted in an important change in perception regarding the use of e-rubrics as an evaluation instrument, as shown by the 28.7% increase in PSTs who felt comfortable with them (from 46.9% before the programme to 75.6% afterwards). The Final Masters Projects were analysed in order to check whether this positive perception of the use of e-rubrics, and the use of GRs, was translated into teaching practice. Figure 9 shows the percentages for transfer of the GR designed into practice, and use of e-rubrics by the PSTs, as obtained upon analysis of their Final Masters Project, for each category.

The results show that 74.4% of PSTs did not manage to transfer the GR designed into teaching practice as it was neither used nor even mentioned in the Final Masters Project. Only 23.3% included their GR as part of the design of the teaching–learning sequence in addition to putting it into practice. When interpreting these findings, we must take into
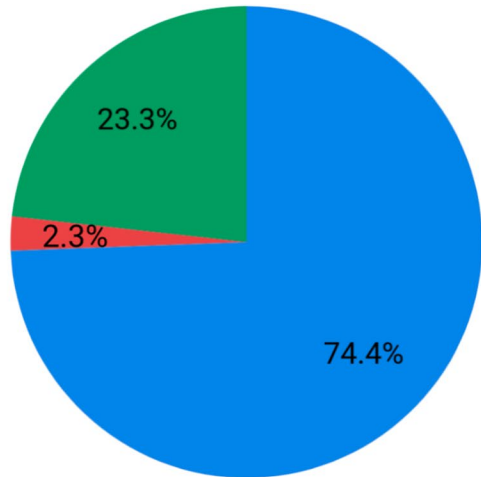
Table 8 Perception of GR and e-rubrics understanding before and after the activity analysed (in percentages)

| Activity	N	Perception of the Understanding	1	2	3	4	5
GRs Training	48	Before activity	**42.6**	38.3	12.8	4.3	2.1
		After activity	0.0	0.0	0.0	33.3	**66.7**
Science Fair	41	Before activity	22.5	**52.5**	20.0	2.5	2.5
		After activity	0.0	0.0	2.5	37.5	**60.0**
E-rubric Training	44	Before activity	**83.7**	11.6	2.3	2.3	0.0
		After activity	0.0	0.0	14.0	39.5	**46.5**
Use of e-rubric in science fair	41	Before activity	25.0	**32.5**	27.5	15.0	0.0
		After activity	0.0	0.0	12.5	**47.5**	40.0
Consensus criteria	44	Before activity	**82.8**	9.8	7.3	0.0	0.0
		After activity	0.0	9.8	**34.1**	**34.1**	22.0

The highest values obtained in each item are shown in bold

1: I do not know anything, 2: I know a little, 3: I know it well, 4: I know it very well, 5: I can explain it to a friend

● 1) GR is not mentioned in the Final Masters Project

● 2) GR is mentioned, but is not part of the teaching-learning sequence activities

● 3) GR is mentioned and is part of the design of activities for the teaching-learning sequence and is put into practice in the intervention

23.3%

2.3%

74.4%

● 1) Rubric is not mentioned in the Final Masters Project

● 2) Rubric is mentioned, but is not part of the teaching-learning sequence design

● 3) Rubric is mentioned and is incorporated into the teaching-learning sequence design but is not involved in the intervention

● 4) Rubric is mentioned, is incorporated into the teaching-learning sequence design and is involved in the intervention, but it is not an e-rubric

● 5) Rubric is mentioned, is incorporated into the teaching-learning sequence design, is involved in the intervention and an e-rubric is used

20.9%

16.3%

14.0%

2.3%

46.5%
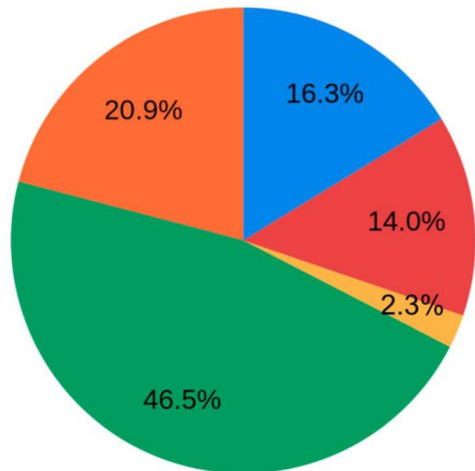
Fig. 9 Categorisation of transfer of the GRs (above) and e-rubrics (below) into practice

account that each GR was designed for a specific topic, which in many cases was impossible to put into practice as the groups of students where the PSTs implemented their teaching–learning sequence may already have covered that topic.

However, there was no such problem with the use of e-rubrics applicable to any topic, with 67.4% of PSTs incorporating the design into their teaching–learning sequence design and implementing a traditional (category 4) or electronic rubric (category 5) into their practice period, with 20.9% of these being electronic. This reflects the importance and awareness of PSTs regarding the use of rubrics as an evaluation instrument, thus allowing a more objective evaluation. It also shows that PSTs feel more able to evaluate using rubrics after the programme.

7 Conclusions and Educational Implications

The results obtained in this study allowed us to draw a series of conclusions regarding the research questions.

The study carried out has shown that the training programme is able to promote reflection in PSTs for the design and improvement of the evaluation criteria for an e-rubric that is to be used to evaluate GRs, as part of a social constructivism approach. Said reflection had an impact on both the evaluation criteria proposed and the selection thereof in a consensual manner (Research Question 1). Thus, the criteria for the e-rubric undergo marked changes during the design and preparation of a GR in parallel, with *adaptation of the GR to the educational level* (M05), the *sources used to prepare the GR* (M08) and *ease of use of the GR* (M13) being the most influential aspects as regards evolution of the criteria. Thus, the evaluation criteria in the e-rubric can be defined adequately during the GR preparation process, although use of the e-rubric itself causes the PSTs to become aware of how the achievement levels for the criteria defined should be scored (Research Question 1a). In addition, the reflections of PSTs after using the e-rubric show that it is more difficult to evaluate a GR than the speaker. This may be due to the fact that more evaluation criteria were identified to evaluate the speaker than the GR.

Similarly, the agreement of consensus criteria by PSTs when preparing an e-rubric results in a more in-depth and significant reflection as regards the criteria that must be applied after evaluation than when these criteria are defined individually, despite a prior reflection concerning them (Research Question 1b). This highlights the importance of a social constructivism that facilitates discussion and reflection amongst PSTs, encouraging them to discuss ideas which they must agree on and support (Akpan & Kennedy, 2020). Thus, the group consensus when preparing e-rubrics is shown to be a key aspect for many PSTs, prompting them to modify their evaluation criteria so that they are closer to a majority agreement that they had previously not even considered, for example in the case of the importance of designing and *adapting the GR to the educational level* (M05) or the *ability of the GR to evaluate understanding* (M18). The selection of this latter criterion for the consensus e-rubric highlights their awareness of the utility of GRs for both the transmission of understanding and as an evaluation instrument.

Moreover, the programme had an impact on the emotions experienced by the PSTs (Research Question 2), which were favourable, especially during preparation and use of the GR, with interest being the most common emotion in all cases. The insecurity regarding use of the e-rubric or dissatisfaction during the consensus, thus suggesting that some PSTs were unwilling to change their criteria, should also be noted.

The programme also had a marked impact on transfer into practice (Research Question 3). This contributed to a positive change in perception with regard to use of the e-rubric as an evaluation instrument (Research Question 3a), as confirmed upon follow-up of its use in the teaching–learning sequences carried out during their practice periods. In contrast, the transfer of GRs was not as successful (23.3% of PSTs) as, in most cases, it was impossible to put them into practice given that the stage of the curriculum which the students had reached did not coincide with the content for which they had been designed (Research Question 3b).

In our opinion, we believe that the training programme presented herein makes a series of contributions to the literature:

1. This programme allows reflection to be promoted during evaluation in the context of gamification and use of e-rubrics, and gives some idea of the type of activities that produce good results in the framework of social constructivism.
2. In contrast to other evaluation methodologies, where student learning is typically dependent on supervision by the tutor when conducting evaluations (Chetcuti & Cutajar, 2014), our programme does not require this and is able to involve students in decision-making when constructing an e-rubric that is perfected several times. This is due to the different reflection activities proposed and the need to choose appropriate criteria for evaluating different GRs with unknown characteristics.
3. The programme also helps to ensure that PSTs both "do science" and "feel science" (Bellocchi et al., 2017), and to that end provides opportunities to reflect on their emotions, such as a science fair where they can share the GRs they have designed or collaborate or exchange ideas to prepare a consensus e-rubric.

According to the results, within the framework of social constructivism and the reflective teacher, the educational implications of this study are:

(a) Involve PSTs in the reflection in complete cycles of design, implementation and evaluation of activities, in this case on GRs and e-rubrics, offering opportunities at different moments of the training programme.

(b) Reflect on the same activity, first individually and then collectively, generating participation processes in group activities. In this sense, the use of science fairs in PSTs, which have proven to promote group reflection, should be further encouraged. In addition, these science fairs could be extended to topics other than GRs, such as inquiry projects (Alarcón et al., 2021).

(c) Encourage group reflection on evaluation instruments to lead to consensus, which will impact the evaluation. Thus, a group consensus to develop e-rubrics can facilitate understanding evaluation criteria that otherwise would not have been considered.

(d) Help PSTs to understand the evaluation criteria when designing an e-rubric. It is essential to keep in mind that an e-rubric containing too many evaluation criteria can be complex. Therefore, a good balance in the number of criteria will have to be found. On the other hand, PSTs are often unaware of the criteria and scope of an e-rubric until it is used, even if they are the authors of the rubric. According to Cebrián-Robles et al., 2014, peer review helps students understand the criteria.

(e) Introduce activities in which PSTs can interact with each other to facilitate bringing out the emotions of PSTs during design, implementation, and evaluation. For instance, science fairs are again a meeting point that brings out these emotions as PSTs interact.

Finally, the main limitation of this study, which has already been noted by O'Keeffe and Paige (2020), is that both time and the opportunity to implement them are required when undertaking reflexive practices in the classroom. In our opinion, and given the results obtained, this is well worth the effort.

Appendix: Emotions Questionnaire

Indicate the emotions you experienced during the following programme activities. You can select more than one emotion.

Emotions	I selected this emotion because…

GRs Training

☐ rejection
☐ concentration
☐ insecurity
☐ interest
☐ boredom
☐ confidence
☐ satisfaction
☐ dissatisfaction
☐ shame

Science Fair

☐ rejection
☐ concentration
☐ insecurity
☐ interest
☐ boredom
☐ confidence
☐ satisfaction
☐ dissatisfaction
☐ shame

E-rubrics Training

☐ rejection
☐ concentration
☐ insecurity
☐ interest
☐ boredom
☐ confidence
☐ satisfaction
☐ dissatisfaction
☐ shame

E-rubric use

☐ rejection
☐ concentration
☐ insecurity
☐ interest
☐ boredom
☐ confidence
☐ satisfaction
☐ dissatisfaction
☐ shame

Consensus rubric

☐ rejection

Emotions		I selected this emotion because…
☐ concentration		
☐ insecurity		
☐ interest		
☐ boredom		
☐ confidence		
☐ satisfaction		
☐ dissatisfaction		
☐ shame		

On a scale of 1 to 5, evaluate your knowledge of each activity before and after performing it.

Activity	Perception of the Understanding	1	2	3	4	5
GRs Training	Before activity					
	After activity					
Science Fair	Before activity					
	After activity					
E-rubric Training	Before activity					
	After activity					
Use of e-rubric in science fair	Before activity					
	After activity					
Consensus criteria	Before activity					
	After activity					

1: I do not know anything, 2: I know a little, 3: I know it well, 4: I know it very well, 5: I can explain it to a friend

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Abell, S. K., & Lederman, N. (2007). *Research on science teacher knowledge*. Routledge.

Akpan, B., & Kennedy, T. J. (2020). Introduction—theory into practice. In B. Akpan & T. J. Kennedy (Eds.), *Science education in theory and practice: An introductory guide to learning theory* (pp. 1–13). Springer.

Alarcón, M.M., Franco-Mariscal, A.J. y Blanco-López, A. (2021). Ferias de experiencias de indagación en ciencias en la formación inicial de maestros de educación infantil. Una oportunidad para la transferencia al aula. In D. Cebrián-Robles, A.J. Franco-Mariscal, T. Lupión, M.C. Acebal and A. Blanco-Lopez (Coords.), Enseñanza de las ciencias y problemas relevantes de la ciudadanía.Transferencia al aula, (pp. 173–187). Barcelona (Spain): Graó.

Annetta, L. (2012). Learning science through video games and simulations. *Science Education, 96*(3), 566–568.

Beauchamp, C. (2015). Reflection in teacher education: Issues emerging from a review of current literature. *Reflective Practice, 16*(1), 123–141.

Bellocchi, A., Quigley, C., & Otrel-Cass, K. (2017). *Exploring emotions, aesthetics and wellbeing in science education Research*. Springer.

Bellocchi, A., Ritchie, S. M., Tobin, K., King, D., Sandhu, M., & Henderson, S. (2014). Emotional climate and high quality learning experiences in science teacher education. *Journal of Research in Science Teaching, 51*(10), 1301–1325.

Blanco, L. J., Guerrero, E., Caballero, A., Brígido, M., & Mellado, V. (2010). The affective dimensions of learning and teaching and teaching mathematics and science. In M. P. Caltone (Ed.), *Handbook of lifelong learning developments* (pp. 265–287). Nova Science Publishers.

Bovill, C. (2020). Co-creation in learning and teaching: The case for a whole-class approach in higher education. *Higher Education, 79*(6), 1023–1037.

Cebrián-de-la-Serna, M., Serrano, J., & Ruiz, M. (2014). eRubrics in cooperative assessment of learning at University. *Comunicar, 21*(43), 153–161.

Cebrián-Robles, D. (2016). CoRubric. Retrieved from https://corubric.com

Cebrián-Robles, D., Serrano, J., & Cebrián-de-la-Serna, M. (2014). Federated eRubric service to facilitate self-regulated learning in the European university model. *European Educational Research Journal, 13*(5), 575–584.

Chetcuti, D., & Cutajar, C. (2014). Implementing peer assessment in a post-secondary (16–18) physics classroom. *International Journal of Science Education, 36*(18), 3101–3124.

Coenders, F., Terlouw, C., Dijkstra, S., & Pieters, J. (2010). The effects of the design and development of a chemistry curriculum reform on teachers' professional growth: A case study. *Journal of Science Teacher Education, 21*(5), 535–557.

Crandall, P. G., Engler, R. K., Beck, D. E., Killian, S. A., O'Bryan, C. A., Jarvis, N., & Clausen, E. (2015). Development of an augmented reality game to teach abstract concepts in food chemistry. *Journal of Food Science Education, 14*(1), 18–23.

Danielowich, R. (2007). Negotiating the conflicts: Reexamining the structure and function of reflection in science teacher learning. *Science Education, 91*(4), 629–663.

De-Barros-Miller, A.M. (2016). Investigating Teachers' Beliefs in the Implementation of Science Inquiry and Science Fair in Three Boston High Schools (Northeastern University). Retrieved from https://repository.library.northeastern.edu/files/neu:cj82q994b/fulltext.pdf

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification." Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, 9–15. New York: MindTrek.

Dörner, R., Göbel, S., Effelsberg, W., & Wiemeyer, J. (Eds.). (2016). *Serious games: Foundations, concepts and practice*. Springer.

Duit, R., & Treagust, D. F. (1998). Learning in science: From behaviourism towards social constructivism and beyond. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 3–25). Kluwer Academic.

Elias, M. J. (2004). The connection between social-emotional learning and learning disabilities: Implications for intervention. *Learning Disability Quarterly, 27*, 53–63.

Folkes, C., & Carmichael, P. (2006). "Learning to assess" and "assessing to learn": The construction of knowledge about assistive technology. *Educational Action Research, 14*(4), 535–545.

Franco-Mariscal, A.J. (2014). Diseño y evaluación del juego didáctico "Química con el mundial de Brasil 2014". [Design and evaluation of the educational game "Chemistry in 2014 Brazil World Cup"]. Educación Química, 25(E1), 276–283.

Franco-Mariscal, A. J., & Cano-Iglesias, M. J. (2011). Elemental B-O-Ne-S. *Journal of Chemical Education, 88*(11), 1551–1552.

Franco-Mariscal, A. J., & Cano-Iglesias, M. J. (2014). Design and implementation of a bingo game for teaching the periodic table. *School Science Review, 95*(353), 56–59.

Franco-Mariscal, A. J., Oliva-Martínez, J. M., & Bernal-Márquez, S. (2012). An educational card game for learning families of chemical elements. *Journal of Chemical Education, 89*(8), 1044–1046.

Fredricks, J. A., Wang, M. T., Schall, J., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction, 43*, 5–15.

Gaydos, M., & Squire, K. (2010). Citizen science: Designing a game for the 21st century. In R. Van-Eck (Ed.), *Interdisciplinary models and tools for serious games: Emerging concepts and future directions* (pp. 289–304). Hershey: Information Science Reference.

Goulette, E., & Swanson, P. (2019). Video Considerations for the world language edTPA. In D. B. A. Mehdi (Ed.), *Advanced methodologies and technologies in modern education delivery* (pp. 948–959). Pennsylvania: IGI Global.

Guillén, F. D., Cabero, J., Llorente, C., & Palacios, A. (2021). Differential analysis of the years of experience of higher education teachers, their digital competence and use of digital resources: Comparative research methods. *Technology, Knowledge and Learning.* https://doi.org/10.1007/s10758-021-09531-4

Haynes, N. M., Ben-Avie, M., & Ensign, J. (Eds.). (2003). *How social and emotional development add up: Getting results in math and science education*. Teachers College Press, Columbia University.

Haz, L., Guarda, T., León, M., Chalén, M., & Arguello, L. (2018). Playability applied in education. In A. Rocha & T. Guarda (Eds.), *Developments and advances in defense and security* (pp. 299–306). Springer.

Iofciu, F., Miron, C., & Antohe, S. (2012). Constructivist approach of evaluation strategies in science education. *Procedia-Social and Behavioral Sciences, 31*, 292–296.

Jiménez-Liso, M. R., Martínez-Chico, M., Avraamidou, L., & López-Gay Lucio-Villegas, R. (2021). Scientific practices in teacher education: The interplay of sense, sensors, and emotions. *Research in Science & Technological Education, 39*(1), 44–67.

Joag, S. D. (2014). An effective method of introducing the periodic table as a crossword puzzle at the high school level. *Journal of Chemical Education, 91*(6), 864–867.

Kagan, J. (2007). What is emotion?: History, measures, and meanings. Binghamton: Vail-Ballou Pr.

Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching, 53*(9), 1316–1340.

Kangas, M. (2010). Creative and playful learning: Learning through game co-creation and games in a playful learning environment. *Thinking Skills and Creativity, 5*(1), 1–15.

Kim, D. (2018). A framework for implementing OER-based lesson design activities for pre-service teachers. *The International Review of Research in Open and Distributed Learning, 19*(4), 149–170.

King, D., Ritchie, S., Sandhu, M., & Henderson, S. (2015). Emotionally intense science activities. *International Journal of Science Education, 37*(12), 1886–1914.

Körkkö, M., Kyrö-Ämmälä, O., & Turunen, T. (2016). Professional development through reflection in teacher education. *Teaching and Teacher Education, 55*, 198–206.

Le Cornu, R., & Peters, J. (2005). Towards constructivist classrooms: The role of the reflective teacher. *Journal of Educational Enquiry, 6*(1), 50–64.

Lee, I. (2007). Preparing pre-service English teachers for reflective practice. *ELT Journal, 61*(4), 321–329.

Lehesvuori, S., Hähkiöniemi, M., Jokiranta, K., Nieminen, P., Hiltunen, J., & Viiri, J. (2017). Enhancing dialogic argumentation in mathematics and science. *Studia Paedagogica, 22*(4), 55–76.

Lei, C., & Chan, C. K. K. (2018). Developing metadiscourse through reflective assessment in knowledge building environments. *Computers & Education, 126*, 153–169.

Linnenbrink, E. A., & Pintrich, P. R. (2002). The role of motivational beliefs in conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 115–135). Kluwer Academic Publishers.

Lu, J., & Zhang, Z. (2012). Understanding the effectiveness of online peer assessment: A path model. *Journal of Educational Computing Research, 46*(3), 313–333.

Luque, E., Franco-Mariscal, A.J. & Blanco, A. (2021). Promoviendo la reflexión y el pensamento crítico en la formación inicial del profesorado de ciencias de educación secundaria. *Actas del XV Congreso Internacional de Educación e Innovación CIEI*. Firenze (Italy). 13–15 december 2021.

Luttikhuizen, P. C. (2018). Teaching evolution using a card game: Negative frequency-dependent selection. *Journal of Biological Education, 52*(2), 122–129.

Marcelo, C. (2009). Professional development of teachers: Past and future. *Sísifo, Educational Sciences Journal, 8*, 5–20.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37*(1), 153–184.

Matthews, B. (2004). Promoting emotional literacy, equity and interest in science lessons for 11–14 year olds; the 'improving science and emotional development' project. *International Journal of Science Education, 26*(3), 281–308.

Matusov, E. (2001). Intersubjectivity as a way of informing teaching design for a community of learners classroom. *Teaching and Teacher Education, 17*(4), 383–402.

McComas, W. F. (2011). Science fair. *The Science Teacher, 78*(8), 34–38.

McKenney, S. (2018). How can the learning sciences (better) impact policy and practice? *Journal of the Learning Sciences, 27*(1), 1–7.

McKinley, J. (2015). Critical argument and writer identity: Social constructivism as a theoretical framework for EFL academic writing. *Critical Inquiry in Language Studies, 12*(3), 184–207.

Moreno, N.M., & Franco-Mariscal, A.J. (2019). Programa formativo de realidad aumentada y realidad virtual en la enseñanza de las ciencias en la educación superior. In A. Alias, D. Cebrián & F. Ruiz (Coords.), *Tecnologías para la formación de profesionales en educación*, (pp. 178–196). Madrid: Dykinson.

Mustafa, G. M. M. (2017). Learning with each other: Peer learning as an academic culture among graduate students in education. *Journal of Educational and Behavioral Statistics, 5*(9), 944–951.

Nicolaou, C. T., Evagorou, M., & Lymbouridou, C. (2015). Elementary school students' emotions when exploring an authentic socio-scientific issue through the use of models. *Science Education International, 26*(2), 240–259.

O'Keeffe, L., & Paige, K. (2020). Reflections on journaling: An initiative to support pre-service mathematics and science teachers. *Australian Journal of Teacher Education, 45*(4), 76–95.

Op't Eynde, P., De Corte, E., & Vershaffe, L. (2006). Accepting emotional complexity: A socio-constructivism perspective on the role of emotions in the mathematics classroom. *Educational Studies in Mathematics, 63*, 193–207.

Parker, J., & Heywood, D. (2013). Exploring how engaging with reflection on learning generates pedagogical insight in science teacher education. *Science Education, 97*(3), 410–441.

Paul, J., Lederman, N. G., & Großa, J. (2016). Learning experimentation through science fairs. *International Journal of Science Education, 38*(15), 2367–2387.

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International handbook of emotions in education*. Routledge.

Perkins, A.C. (2016). Earthquake: Game-based learning for 21st century STEM education (Texas A&M University). Retrieved from http://hdl.handle.net/1969.1/157955

Perrenoud, P. (2004). Diez nuevas competencias para enseñar. [Ten new competences to teach]. Barcelona: Graó.

Poerksen, B. (2004). 'We can never know what goes on in somebody else's head': Ernst von Glasersfeld on truth and viability, language and knowledge, and the premises of constructivist education. *Cybernetics and Systems, 35*(4), 379–398.

Polin, L. (2018). A Constructivist Perspective on Games in Education. In D. W. Kritt (Ed.), *Constructivist education in an age of accountability*. Springer: Palgrave.

Quintana, A. M. V., Rogado, A. B. G., Gavilán, A. B. R., Martín, I. R., Esteban, M. A. R., Zorrilla, T. A., & Izard, J. F. M. (2014). Application of new assessment tools in engineering studies: The rubric. *IEEE Revista Iberoamericana De Tecnologías Del Aprendizaje, 9*(4), 139–143.

Rannikmäe, M., Holbrook, J., & Soobard, R. (2020). Social constructivism-jerome bruner. In B. Akpan & T. J. Kennedy (Eds.), *Science education in theory and practice*. Cham: Springer.

Ritchie, S. M., Tobin, K., Hudson, P., Roth, W. M., & Mergard, V. (2011). Reproducing successful rituals in bad times: Exploring emotional interactions of a new science teacher. *Science Education, 95*(4), 745–765.

Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*, 247–270.

Russell, T. (2005). Can reflective practice be taught? *Reflective Practice, 6*, 199–204.

Russell, T. (2012). Paradigmatic changes in teacher education: The perils, pitfalls, and unrealized promise of the reflective practitioner. *Encounters in Theory and History of Education, 13*, 71–91.

Ryan, T. G., Young, D. C., & Kraglund, W. L. (2017). Action research within pre-service teacher education. *Transformative Dialogues: Teaching & Learning Journal, 10*(3), 1–18.

Sahin, A. (2013). stem clubs and science fair competitions: Effects on post-secondary matriculation. *Journal of STEM Education: Innovations and Research, 14*(1), 5–11.

Schön, D. A. (1992). *The reflective practitioner: How professionals think in action*. Basic Books.

Sebre, S. B., & Miltuze, A. (2021). Digital media as a medium for adolescent identity development. *Technology, Knowledge and Learning, 26*(4), 867–881.

Shoffner, M. (2009). The place of the personal: Exploring the affective domain through reflection in teacher preparation. *Teaching and Teacher Education, 25*(6), 783–789.

Sinatra, G. L., Broughton, S. H., & Lombardi, D. (2014). Emotions in Science education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 415–437). Routledge.

Slepcevic-Zach, P., & Stock, M. (2018). ePortfolio as a tool for reflection and self-reflection. *Reflective Practice, 19*(3), 291–307.

Smith, P., Geng, G., & Black, P. (2017). Teachers as Reflective Practitioners. In G. Geng, P. Smith, & P. Black (Eds.), The challenge of teaching. Through the eyes of pre-service teachers (pp. 25–32). Singapore: Springer.

Steffens, K., & Underwood, J. (2008). Self-regulated learning in a digital world. *Technology, Pedagogy and Education, 17*(3), 167–170.

Stevens, D. D., & Levi, A. J. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Stylus Publishing LLC.

Stieglitz, S., Lattemann, C., Robra-Bissantz, S., Zarnekow, R., & Brockmann, T. (Eds.). (2017). *Gamification. Using game elements in serious contexts*. Springer.

Tan, E., Calabrese, A., & Benavides, A. (2019). Engineering for sustainable communities: Epistemic tools in support of equitable and consequential middle school engineering. *Science Education, 103*(4), 1011–1046.

Tee, D. D., & Ahmed, P. K. (2014). 360 degree feedback: An integrative framework for learning and assessment. *Teaching in Higher Education, 19*(6), 579–591.

Van Beveren, L., Roets, G., Buysse, A., & Rutten, K. (2018). We all reflect, but why? A systematic review of the purposes of reflection in higher education in social and behavioral sciences. *Educational Research Review, 24*, 1–9.

Wanner, T., & Palmer, E. (2018). Formative self-and peer assessment for improved student learning: The crucial factors of design, teacher participation and feedback. *Assessment & Evaluation in Higher Education, 43*(7), 1032–1047.

Woolfolk, A. (2014). *Educational psychology*. Noida, India: Dorling Kindersley India Pvt. Ltd.

Yun, E., & Park, Y. (2018). Extraction of scientific semantic networks from science textbooks and comparison with science teachers' spoken language by text network analysis. *International Journal of Science Education, 40*(7), 2118–2136.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147.