Acta Scientiarum

HISTÓRY AND PHILOSOPHY OF EDUCATION

Novice teachers in the face of assessment: analysis of learning situations in the MIR for teachers in the Community of Madrid

Noelia Santamaría-Cárdaba*, **Vanessa Ortega-Quevedo and Judith Cáceres-Iglesias**

Universidad de Valladolid, Campus María Zambrano, Pl. de la Universidad, 1, 40005, Segovia, Espanha. *Author for correspondence: E-mail: noelia.santamaria.cardaba@uva.es

ABSTRACT. Assessment is an issue present in every educational process, so this study focuses on analyzing the treatment of assessment in the Learning Situations designed by novice teachers during the Comprehensive Teacher Training program (hereinafter, educational MIR) of the Community of Madrid. The methodology employed was mixed, using quantitative and qualitative analyses to study the information on assessment available in the 390 Learning Situations analyzed. The results reveal that heteroassessment continues to predominate among teachers and that there are inconsistencies between the assessment tools and the assessment process they propose in their Learning Situations. In conclusion, novice teachers show a lack of training to correctly apply a formative and shared assessment model with their students.

Keywords: assessment; learning situations; educational MIR; novice teacher.

Los docentes noveles ante la evaluación: análisis de situaciones de aprendizaje en el MIR de maestros de la Comunidad de Madrid

RESUMEN. La evaluación es una cuestión presente en todo proceso educativo, por ello este estudio se centra en analizar el tratamiento de esta en las Situaciones de Aprendizaje diseñadas por los docentes noveles durante el programa de Capacitación Integral Docente (en lo sucesivo, MIR educativo) de la Comunidad de Madrid. La metodología empleada ha sido mixta empleando análisis cuantitativos y cualitativos para estudiar en profundidad la información sobre evaluación disponible en las 390 Situaciones de Aprendizaje analizadas. Los resultados revelan que la heteroevaluación sigue predominando entre el profesorado y que existen incoherencias entre las herramientas de evaluación y el proceso evaluativo que plantean en sus Situaciones de Aprendizaje. En conclusión, el profesorado novel muestra una carencia formativa para aplicar correctamente un modelo de evaluación formativa y compartida con su alumnado.

Palabras clave: evaluación; situaciones de aprendizaje; MIR educativo; profesorado novel.

Novos professores diante da avaliação: análise de situações de aprendizagem no MIR para professores da Comunidade de Madrid

RESUMO. A avaliação é uma questão presente em todos os processos educativos, pelo que este estudo se centra na análise do tratamento da avaliação nas Situações de Aprendizagem concebidas por professores principiantes durante o programa de Formação Integral de Professores (doravante designado por MIR educativo) na Comunidade de Madrid. A metodologia utilizada foi mista, empregando análises quantitativas e qualitativas para estudar em profundidade as informações sobre avaliação disponíveis nas 390 Situações de Aprendizagem analisadas. Os resultados revelam que a heteroavaliação continua predominando entre os professores e que há inconsistências entre os instrumentos de avaliação e o processo de avaliação que eles propõem em suas Situações de Aprendizagem. Em conclusão, os professores iniciantes demonstram falta de treinamento para aplicar corretamente um modelo de avaliação formativa e compartilhada com seus alunos.

Palavras-chave: avaliação; situações de aprendizagem; MIR educacional; novos professores.

Introduction

Assessment in education is a particularly relevant topic, since '[...] it is impossible to conceive and organise a teaching-learning process without the assessment component' (Ocampo, 2017, p. 4). In view of this, this study focuses on analysing the assessment designed in the Learning Situations (hereinafter, LS) developed by new teachers enrolled in the Comprehensive Teacher Training programme, known as MIR Educativo, in the Community of Madrid.

Authors such as Egido Gálvez (2021) and López-Rupérez (2018, 2021) understand this educational MIR as '[...] a tool for strengthening the teaching profession and, therefore, contributing to the improvement of the education system' (Egido Gálvez, 2021, p. 211). Specifically, the Comprehensive Teacher Training programme discussed in this study is an initiative implemented for the first time in the 2022/2023 academic year in Madrid schools, in which teachers, before joining the civil service, complete a one-year internship to further their training. During this period, they undergo a final assessment that definitively accredits them as civil servants. In addition, this educational MIR involves a five-module training programme in which they can access '[...] videos, consultation forums, webinars with experts, practical exercises and case studies to help them take up their duties with confidence, responsibility and communication and leadership skills' (Community of Madrid, 2022).

One of the mandatory tasks of the educational MIR (Medical Intern Residency) programme in the Community of Madrid is to develop a learning situation using innovative methodologies and detailing the assessment they intend to carry out during the teaching-learning process. This is where the material analysed in this study comes in, as we focus on studying the educational assessment proposed in these LS. Now, what is a Learning Situation? Parra (2013) defines LS as a set of learning activities and complex tasks designed to achieve established objectives. With regard to assessment, he highlights the importance of it being part of learning. He does not refer to LS in conjunction with assessment, but rather that assessment is integrated into the learning process, which ties in with the views of Richard and Godbout (2000). In this way, students do not learn in order to be assessed, but are assessed in order to learn.

This is in line with Royal Decree No. 157 (2022), which establishes the organisation and minimum teaching requirements for primary education within the framework of the LOMLOE. Article 2 includes, among other definitions, that of SA, which is understood as: 'Situations and activities that involve the deployment by students of actions associated with key competences and specific competences and that contribute to the acquisition and development of these' (Royal Decree No. 157, 2022, p. 24388).

With regard to assessment, this Royal Decree advocates continuous, formative and inclusive assessment. In line with this proposal and as it is a key element in the development of this research, we understand this assessment process as set out and defended by the Formative and Shared Assessment Network, which in its studies (López-Pastor & Pérez-Pueyo, 2017; López-Pastor et al., 2016) defines formative assessment as any assessment process that aims to improve the teaching-learning process in three areas: improving student learning; improving the teacher's teaching process; and improving the teaching-learning process itself.

Consequently, all teaching and learning activities must establish an assessment process that allows decisions to be made and the teaching and learning process to be improved. However, not all activities that make up a teaching and learning activity need to be graded (Calvo et al., 2024). This is something that needs to be made clear from the theoretical concepts relating to assessment in teaching and learning activities, as in Spain the confusion between grading and assessment seems to be deeply rooted. The notion of "continuous assessment" in this country has been interpreted by most teachers as synonymous with "continuous exams and grading", repeatedly making the same mistake, instead of doing so consistently and formatively throughout the school year (Álvarez-Méndez, 1993, 2000, 2003; Fernández-Pérez, 1986; López-Pastor & Pérez-Pueyo, 2017; Santos Guerra, 1993, 2003).

Various studies, such as those by Dochy et al. (2002), Fernández-Pérez (1986, 1989), Fernández-Sierra (1996) and López-Pastor and Pérez-Pueyo (2017), have found that there is confusion among teachers who confuse the concept of "assessment" with "giving marks", i.e. the perception of assessment is reduced to grading. Along these lines, authors such as Santos Guerra (2003) and Rogero (2015, p. 10) argue that "[...] assessment has the meaning it has: to classify, grade, disqualify, select, exclude, label, tame, subdue, docilise... and nothing else". However, this idea of assessment as grading is very limited and does not fit with the proposed formative assessment model, as formative assessment is distinguished by its contribution to the teaching-learning process and not by assigning a grade (Marshall & Drummond, 2006). Likewise, it has been

proven that assessment without grading favours the teaching-learning process of students (López-Pastor & Pérez-Pueyo, 2017).

The complexity of the term "assessment" has led to the emergence of various studies focused on researching assessment in education in Spain (Álvarez-Méndez, 2000, 2003; Ibarra-Sáiz & Rodríguez-Gómez, 2015; López-Pastor, 2006, 2010) and, years earlier, at the international level (Antoniou & James, 2014; Higgins, 2013; López-Pastor & Pérez-Pueyo, 2017; Stake, 1967; Wiliam et al., 2004). These studies are based on the review carried out in 1998 by Black and Wiliam (1998) to develop the state of the art, which shows that internationally there is already talk of formative assessment, learning-oriented assessment or integrated assessment, while in Spain the classic concepts of summative assessment, self-assessment or final assessment are still gaining ground.

This scenario reflects that the concept of educational assessment is a dilemma (Poggi, 2008) and that new ways of conceptualising assessment arise from the need to understand and develop the assessment process in a different way. Nevertheless, in line with López-Pastor and Pérez-Pueyo (2017), in this study we understand that all these approaches are encompassed in the concept of "formative assessment", understood as:

> Any assessment process whose main purpose is to improve the teaching-learning processes that take place. It helps students learn more (and/or correct their mistakes) and teachers learn to work better (to improve their teaching practice). In other words, the main purpose is not to grade the student, but to obtain information that allows us to know how to help students improve and learn more, and which in turn helps teachers learn to do their job better and better (López-Pastor, 2007, p. 64).

It should be noted that assessment in the SA that will be analysed is defined by each teacher, so they can implement different strategies such as those highlighted by López-Pastor et al. (2005) and Basurto-Mendoza et al. (2021): 1) self-assessment, understood as the assessment that a person makes of themselves, whether by the teacher or the students; 2) co-evaluation, which is carried out between peers, i.e. between the students themselves; or 3) hetero-evaluation, understood as the assessment made by one person of another who is not their peer, which can be teacher-student (the most common) or student-teacher. Similarly, according to Hamodi et al. (2015), different tools can be used to carry out these assessments, such as: teacher's diary, numerical scales, descriptive rubrics, observation sheets, checklists, self-assessment and peer assessment sheets, etc.

Subject of study

In line with the above, this research focuses on studying how formative assessment, which must be integrated into the SA programmed in the final activity of the educational MIR, is reflected. Specifically, the research objectives are as follows:
- To analyse what types of assessment new teachers consider in the programmed SA, checking whether the formative assessment required by the SA is used.
- Discover what types of assessment tools teachers use in their SAs.
- Check the consistency in the use of assessment instruments in the SA.

Methodology

This article uses a mixed methodology, which '[...] involves a process of collecting, analysing and linking quantitative and qualitative data in the same study' (Hernández Sampieri et al, 2006, p. 751). It should be added that applying this mixed methodology to address the objectives of this study helps to enhance '[...] the research process in education' (Bagur-Pons et al., 2021, p. 16), which is the area on which our research focuses.

Participants

This research focuses on the study of 527 teaching plans programmed and delivered in Spanish by novice teachers who were studying for the MIR educational qualification in the Community of Madrid. Specifically, the novice teachers delivered the presentation associated with the programming of a teaching plan on a free topic that had been assigned to them 15 days in advance.

From these 527 documents, a sample of 390 was selected at random for analysis in this study. Therefore, based on the size of the population and the sample, the data analysed with a 95% confidence level has only a 5% margin of error. With regard to the teachers who programmed the 390 SA designed, it should be noted that 278 are women, 56 are men, and in 56 cases no data has been identified to show the gender of the participants.

Quantitative analysis procedure

The quantitative analysis data is derived from a prior inductive analysis of the SA. To this end, the researchers created an Excel template where variables (educational area, educational level, types and instruments of assessment) were counted and recorded. After filling in the data in Excel, the document was imported into the SPSS IBM Statistics 29 programme, from which two types of analysis were performed: 1) descriptive, which allows basic frequencies to be obtained for easy reading of the results (Hidalgo Troya, 2019); and 2) inferential, based on comparison using cross tables and Pearson's X2 test.

Qualitative analysis procedure

The quantitative study of the data has been enriched in this research with a qualitative analysis of the same. This qualitative analysis helps to deepen the research results given that '[...] it studies the association or relationship between the variables that have been quantified, which further aids in the interpretation of the results' (Sarduy Domínguez, 2007, p. 5). After a preliminary general reading of the information, a deductive coding of the data was carried out to perform a qualitative analysis for thematic robustness; along these lines, the categories and subcategories of analysis shown in Table 1 were established.

Table 1. Categories and subcategories of qualitative analysis.

Categories	Subcategories
Dimensions	- Type of assessment - Techniques and tools: what I use to assess (rubrics, guidelines, etc.) - Method: how I do it (routines, cooperatives) - Grading
Inconsistencies	- Terminological confusion - Type of assessment inconsistent with objects and/or techniques

Source: Own work.

Results

Quantitative results

The results derived from the descriptive analysis are presented below. First, the results obtained for the courses in which teachers who have completed the educational MIR have decided to schedule their SA are shown. As we can see in Table 2, the course in which most teachers decide to schedule is the 5th year, and the course in which the fewest schedules have been recorded is the 2nd year. If we analyse the results by cycle, we see that the cycle in which the fewest SAs have been programmed is the first cycle, and the cycle in which the most have been programmed is the third (frequency 62, 146 and 169 respectively). Furthermore, we can see that most teachers prefer to programme in the first years of the cycle (1st, 3rd and 5th), with the last years having the lowest number of SAs.

Table 2. SA frequency per course.

	1º	2º	3º	4º	5º	6º	Lost	Total
Frecuency	36	26	91	55	124	45	13	390
%	9,2	6,7	23,3	14,1	31,8	11,5	3,3	100

Source: Own work.

By studying the relationship between the course in which it is programmed and the rest of the variables using the χ^2 test, we can see that there are significant differences in terms of teacher-student self-assessment (p value = .032). Upon in-depth analysis of this relationship, we can observe (Table 3) that in the first cycle courses, the relationship between teachers who use student self-assessment and those who do not is more balanced; although more teachers use it than do not, the frequency is similar. In the second cycle, there is a discrepancy between the courses, as there is greater use of student self-assessment in the SA programmed for the third year than for the fourth year. Furthermore, the fourth year is the only year for which more SA have been designed that do not contain self-assessment compared to those that do. In the third cycle, we find more SAs with self-assessment than those that do not use this assessment, with it being more prevalent in the fifth year than in the sixth year.

Secondly, we move on to study the SA programmed by main area. According to Table 4, the main area in which teachers programmed most frequently is Spanish Language and Literature, followed by Mathematics.

In third place is Natural Sciences (CCNN) and in fourth place is Social Sciences (CCSS), with an anecdotal appearance for Social and Civic Values and Art Education.

Table 3. SA scheduled by course in relation to the completion of self-assessment by students.

F lost		F 1º		F 2º		F3º		F 4º		F 5º		F 6º	
No A	Yes A	No A	Yes A	No A	Yes A	No A	Yes A	No A	Yes A	No A	Yes A	No A	Yes A
11	2	16	20	12	14	38	53	31	24	42	98	12	33

Note: No A = No self-assessment carried out; Yes A = Self-assessment carried out. Source: Own elaboration.

Table 4. SAs scheduled by main area.

	Mathematics	Spanish Language	CCNN	CCSS	Values	Art Education	Lost	Total
Frecuency	114	147	80	39	2	1	7	390
%	29,2	37,7	20,5	10	,5	,3	1,8	100

Source: Own elaboration.

When contrasting this variable using the χ^2 test, several significant differences are detected. The first is recorded in the area of Spanish Language and Literature, where there are significant differences regarding the inclusion of self-assessment (p = .003). Specifically, when studying the case, clear differences are observed in the number of SAs that include self-assessment in this area (100 SAs with self-assessment) in contrast to those that do not include it (47 SAs that do not contemplate student self-assessment). This same relationship, but in reverse, is found in the area of mathematics (p = .03), where there are more teaching plans that do not include self-assessment (276) than those that do (114).

The second significant relationship is observed between the planned teaching and learning activities in the area of Natural Sciences and teacher self-assessment. On this occasion, the same relationship is identified as in the area of language, with teaching and learning activities involving self-assessment being more frequent in this area. Specifically, there were 52 teaching and learning activities involving student self-assessment compared to 28 without it (p = 0.043).

Thirdly, the types of assessment selected by teachers in the SA developed during the MIR are studied. These types are teacher-student hetero-assessment (HM-A), co-assessment or peer assessment (CE), student self-assessment (AA), teacher self-assessment and/or Learning Situation self-assessment (AMSA), and student-teacher hetero-assessment (HA-D). Table 5 shows that the most common type of assessment is hetero-assessment, followed by teacher and/or learning situation self-assessment and student self-assessment. It is noteworthy that only 44.4% of the SA include co-evaluation processes and that only 10.5% of these include student-teacher hetero-evaluation processes.

Table 5. Frequency of use of different types of assessments in SA.

	HM-A		CE		AA		AMSA		HA-D	
	No	Sí	No	Sí	No	Sí	No	Sí	No	Sí
Frecuency	12*	378	217	173	162	228	69	321	349	41
%	3,1	96,9	55,6	44,4	41,5	58,5	17,7	82,3	89,5	10,5

*Note: these 12 cases are considered lost because it is not that they do not use hetero-evaluation, but rather that it is not manifested as such. Source: Own elaboration.

Regarding the relationship between the types of assessment using Pearson's chi-square test, the following relationships are noteworthy. Firstly, the analysis of the relationships between student self-assessment and other types of assessment shows significance in the relationships with peer assessment (HM-A) (p < 0.001), co-assessment (p < 0.001) and teacher self-assessment (AMSA) (p = 0.005). In the first case, although the relationship is significant, it does not provide much information, since, except for the missing cases, all teachers perform HM-A, so the figures stand out, although their interpretation establishes them as something common.

In the second case, we find the results presented in Table 6, which shows that most teachers who carry out student self-assessment (SSA) also take into account other types of assessment such as peer assessment (PA) and teacher self-assessment (TSA). However, we see that this relationship is more pronounced in the relationship between teacher self-assessment (TSA) and student self-assessment (SSA), since only 30 teachers who choose to conduct student self-assessment have not chosen to self-assess themselves or the SA designed.

Table 6. Relationship between the frequency of AA use in contrast to the frequency of CE and AMSA use.

AA	Coevaluation		AMSA	
	No CE	Yes CE	No AMSA	AMSA
No AA	130	32	39	123
Yes AA	87	162	30	198

Source: Own elaboration.

Finally, the analysis of co-evaluation in relation to the other types of evaluation is shown (Table 7). In this case, only one significant result is obtained when relating co-evaluation (CE) to student-teacher hetero-evaluation (HA-M) ($p = 0.009$). When studying this case, we must remember that only 41 teachers have included this type of assessment in their AS, and within such a low N, the data show that among the teachers who do include HA-M, the majority also include CE.

Table 7. Relationship between the frequency of co-evaluation and the frequency of student-teacher hetero-evaluation.

	No HA-M	Yes HA-M
No CE	202	15
Yes CE	147	26

Source: Own elaboration.

Qualitative analysis

This section details the qualitative analysis organised according to the categories and subcategories that emerged from deductive coding (Table 1). To this end, examples of the SAs that appear are cited as SA, followed by the number corresponding to the order of analysis (e.g. SA1). This section details the qualitative analysis organised according to the categories and subcategories that emerged from deductive coding (Table 1). To this end, examples of the SAs that appear are cited as SA, followed by the number corresponding to the order of analysis (e.g. SA1).

Dimensions

This section describes the dimensions that make up our object of study: the assessment proposals made by teachers enrolled in the educational MIR programme. To this end, we will delve into the type of assessment, what they use to assess (instrument), how they do it (methodology) and grading.

Type of assessment

With regard to the type of assessment, it has been observed that most teachers tend to opt for: student self-assessment, peer assessment, teacher-student assessment, teacher self-assessment, student-teacher assessment, or to propose another form of assessment by selecting the 'other' option. It should be clarified that it is possible to choose one or more types of assessment. In this regard, some teachers choose to select only one type of assessment (D7), but the vast majority select more than one (SA153, SA174). In this selection, all teachers, without exception, choose teacher-to-student assessment (SA12, SA14, SA78, SA106), while very few select student-to-teacher assessment (SA11, SA391).

On the other hand, although some teachers do not select any type of self-assessment (SA54, SA158), the most common choice is teacher self-assessment (SA5, SA17), followed by student self-assessment (SA42, SA191). With regard to the possibility of proposing another form of assessment, the most common option is to resort to continuous assessment (SA54, SA156).

Finally, peer assessment is rarely selected among teachers (SA19, SA391) and, as will be detailed in the 'inconsistencies' category, some of those who choose this option do so incorrectly.

Techniques and instruments

Some SAs have a very comprehensive section on techniques and instruments (SA15, SA23, SA331, SA384) used for assessment. They present a wide variety of instruments, such as rubrics, templates, SWOT analyses, daily anecdotes, checklists, student notebooks, targets, scales, written and oral tests, among others. It should be noted that the most commonly used techniques and instruments are rubrics (SA16, SA369), estimation scales (SA46, SA75) or Likert scales (SA52, SA73), co-evaluation (SA64, SA370) and written tests (SA103, SA145).

Oral tests, also known as oral presentations, are frequently mentioned, especially by teachers who present an SA in the area of Spanish Language (SA6, SA175), although they are also used in SAs in other areas, especially in mathematics, which relate the SA to the area of Spanish Language (SA151, SA389).

Methodology

One of the most widely used methods is peer assessment, which is usually associated with evaluating attitude, behaviour and/or teamwork (SA3, SA42). Teamwork (SA230, SA348), thinking routines (SA1, SA73) and cooperative work (SA333, SA381) are also commonly mentioned and often appear in content assessment. With regard to the latter, expert groups (SA333, SA381), rotating folios (SA377, SA384), followed by expert minds (SA381) or pencils in the centre (SA377), are the most common activities.

Rating

Two groups can be distinguished in terms of grading. On the one hand, there are those who prefer not to specify how the grading process will be carried out and develop the assessment process extensively (SA37, SA280). On the other hand, there are those who opt for a grading system based on continuous, formative and summative assessment. For this reason, in the grading section, they detail the percentages that will be taken into account for the final grade (SA29, SA384). Among the most common are percentages for: final product, individual tasks, group tasks, oral presentation, participation and attitude.

Inconsistencies

Terminological confusion

Peer assessment is one of the most frequently cited assessment methods (SA249, SA65), but none of these citations mention external reviewers such as families, teachers of other subjects, etc. Nor is there any mention of student-teacher assessment in most cases. These data indicate that peer assessment is understood as the most traditional form of assessment: teacher-student assessment.

On the other hand, there is terminological confusion regarding the instruments used for the assessment process, SA321 being an example of this. This Learning Situation describes co-assessment using rubrics, but the instrument used is a graded, non-descriptive scale.

As we pointed out in the previous section, the SA analysed do not always explain the grade. This fact shows that teachers have begun to separate these concepts and there are SA with complete assessment processes that include different processes and instruments but do not reflect the jump to the grade (SA176, SA210). However, others maintain this confusion between concepts and link assessment with summative grades (SA45, SA211), which leads them to a final grade for the Learning Situation.

Type of assessment inconsistent with objects and/or techniques

SA11 does not mention student self-assessment as a method; however, in the assessment objectives, it explains that students assess the project they have carried out, evaluating themselves and their peers. This shows that there is an inconsistency between the assessment objectives and the chosen method.

Similarly, SA1 mentions that a student-teacher assessment is carried out, but this process is not reflected in the assessment objectives (initial registration form with a thinking routine, cooperative groups to complete the initial form, individual assessment of an activity) or in the techniques and tools (initial thinking routine, radio programme and self-assessment).

For its part, SA7 selects as its assessment tool or technique: the rubric for assessing SA and self-assessment for students and teachers. The only subject of study it mentions is the assessment of SA, and for grading it opts for hetero-assessment using a rubric. As can be seen, although different methods and techniques are mentioned, grading is summarised in a hetero-assessment that is not explained how it is carried out. Thus, the assessment process proposed with the grading is inconsistent, as the self-assessment by students or teachers is not reflected in the grading in any way.

Discussion

The results detailed above show that when designing assessment processes, teachers must make numerous decisions about a wide range of curricular elements (Bearman et al., 2016; Ibarra-Saiz et al., 2023). These decisions include the types of assessment they wish to carry out and the instruments they intend to use to assess SA.

The data from this study show that hetero-assessment is the form of assessment present in all AS, which means that teacher-student assessment continues to predominate among novice teachers. In this regard, authors such as Bilbao Martinez and Villa Sánchez (2019), Gallardo-Fuentes et al. (2017) and Hernández-Infante and Infante-Miranda (2017) found that in universities, and specifically in early childhood and primary education degrees, the predominant form of assessment is hetero-assessment. Thus, our results reveal that teachers who have recently graduated from education faculties reflect the assessment model they experienced during their university studies in their own teaching practices.

Student participation in assessment is generally achieved through the use of co-assessment or peer assessment, while incorporating student-teacher hetero-assessment is not common in the designed SAs. With regard to co-assessment, López-Pastor et al. (2005) already pointed out the importance of getting students to participate in their assessment in order to improve their training through techniques such as co-assessment. Studies such as those by Sanmartí Puig (2019) and Torregrosa Martinez et al. (2021, p. 109) have even found that 'within the problem-solving process [...] co-evaluation acts as an effective tool for sharing different methods of resolution.' For their part, teacher evaluation is addressed by authors such as Prats et al. (2020), García Garduño and Medécigo Shej (2014) and Villa Sanchez (1985), who agree that this type of assessment helps teachers to learn and improve their educational practice.

It should be noted that the assessment tools most frequently used by teachers in their SA are rubrics, estimation or grading scales, and written tests. These data reveal that exams continue to be one of the assessment techniques used by teachers, despite studies such as those by Torrano-Martínez et al. (2017) and Seijo Galán et al. (2020) highlighting the anxiety that this type of test causes in students. A formative and shared assessment would avoid this type of situation, which would involve using other assessment tools such as thinking routines or portfolios (Barba-Martín et al., 2020; Hamodi et al., 2015). Along these lines, another aspect to highlight is the inconsistencies noted. The results reveal inconsistencies arising from terminological confusion about assessment techniques and contradictions between the proposed assessment process and the grading, which coincides with the findings of Iglesia Villasol (2020), López-Pastor et al. (2008) and Zaragoza Costerad et al. (2009).

All of this shows that some of the SA programmed by new teachers do not use comprehensive formative assessment as understood by the Formative Assessment Network. We see some cases where there are signs of this deep-rooted culture of grading and examination, and the focus on improving student learning, the teacher's teaching process and the teaching-learning process itself is not so clear.

Conclusion

This study leads to the conclusion that teachers enrolled in the MIR programme in the Community of Madrid have reflected in the programmed SA a way of understanding assessment centred on hetero-assessment. In some cases, the concept of assessment continues to be confused with percentages and grades, and trends such as the use of written tests as an assessment tool persist. This reveals the need to improve their training in order to promote formative and shared assessment in their teaching practice.

Finally, it is worth noting some limitations and future lines of research. The main limitation is considered to be access only to the presentations of the SA, without having contact with the teachers who produce them to learn more about the reasons for the assessment system they present, how they apply it, or the context of the group for which it is designed. These limitations can be understood as future lines of study, seeking in the next educational MIR to have closer contact with teachers in order to understand their decisions and detect possible training needs.

Data availability

Not applicable.

References

Álvarez-Méndez, J. M. (1993). El alumnado: la evaluación como actividad crítica de aprendizaje. *Cuadernos Pedagogí*a, *219*, 28-32.

Álvarez-Méndez, J. M. (2000). *Evaluar para conocer, examinar para excluir*. Morata.

Álvarez-Méndez, J. M. (2003). *La evaluación a examen. Ensayos críticos*. Miño y Dávila.

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability, 26*, 153-176. https://doi.org/10.1007/s11092-013-9188-4

Bagur Pons, S., Rosselló-Ramon, M. R., Paz-Lourido, B., & Verger Gelabert, S. (2021). El enfoque integrador de la metodología mixta en la investigación educativa. *RELIEVE. Investigación y Evaluación Educativa, 27*(1), 1-21. http://doi.org/10.30827/relieve.v27i1.21053

Barba-Martín, R., Hortigüela Alcalá, D., & Pérez-Pueyo, Á. (2020). Evaluar en educación física: análisis de la tensiones existentes y justificación del empleo de la evaluación formativa y compartida. *Educación Física y Deporte, 39*(1), 23-46. https://doi.org/10.17533/udea.efyd.v39n1a03

Basurto-Mendoza, S. T., Moreira-Cedeño, J. A., Velásques-Espinales, A. N., & Rodríguez-Gámez, M. (2021). Autoevaluación, coevaluación y heteroevaluación como enfoque innovador en la práctica pedagógica y su efecto en el proceso de enseñanza-aprendizaje. *Polo del Conocimiento: Revista Científico-Profesional, 6*(3), 828-845. https://doi.org/10.23857/pc.v6i3.2408

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: developing the assessment design decisions framework. *Teaching in Higher Education, 21*(5), 545-556. https://doi.org/10.1080/13562517.2016.1160217

Bilbao Martinez, A., & Villa Sánchez, A. (2019). Avances y limitaciones en la evaluación del aprendizaje a partir del proceso de convergencia. Visión docente y discente en los grados de Educación Infantil y Primaria. *Educación XX1, 22*(1), 45-69. https://doi.org/10.5944/educxx1.19976

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Calvo, A., Castro, A., Haya, I., Lázaro, S., Moral, C., Voces, J., Fernández, C., González, I., Herreros, I., Pardo, J., Sánchez, A., Mier, J. Á., García, M. C., Cuesta, Á., Martínes, A., Salcines, M., Tordable, V., Torre, N., Álvarez, P., … Larrucea, G. U. (2024). *Diseño de SA: de la definición a la práctica*. Universidad de Cantabria.

Comunidad de Madrid (2022, June 8). La Comunidad de Madrid concluye la formación de los primeros tutores de docentes en prácticas que se incorporarán en septiembre a los centros públicos. https://www.comunidad.madrid/noticias/2022/06/08/comunidad-madrid-concluye-formacion-primeros-tutores-docentes-practicas-incorporaran-septiembre-centros-publicos

Dochy, F., Segers, M., & Dierick, S. (2002). Nuevas vías de aprendizaje y enseñanza y sus consecuencias: una era de evaluación. *Revista de Docencia Universitaria, 2*(2), 13-30.

Egido Gálvez, I. (2021). Los modelos médicos aplicados al profesorado: la propuesta del "MIR educativo" a la luz de las experiencias internacionales de iniciación a la profesión docente. *Revista de Educación, 393*, 207-229. https://doi.org/0.4438/1988-592X-RE-2021-393-491

Fernández-Pérez, M. A. (1986) *Evaluación y cambio educativo: análisis cualitativo del fracaso escolar*. Morata.

Fernández-Pérez, M. A. (1989). *Así enseña nuestra universidad*: hacia la construcción crítica de una didáctica universitaria. Hispagraphis.

Fernández-Sierra, J. (1996). ¿Evaluación? No, gracias, calificación. *Cuadernos de Pedagogía, 243*, 92-97. https://dialnet.unirioja.es/metricas/documentos/ARTREV/36044

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

García Garduño, J. M., & Medécigo Shej, A. (2014). Los criterios que emplean los estudiantes universitarios para evaluar la in-eficacia docente de sus profesores. *Perfiles educativos, 36*(143), 124-139. https://doi.org/10.22201/iisue.24486167e.2014.143.44026

Hamodi, C., López Pastor, V. M., & López Pastor, A. T. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida del aprendizaje en educación superior. *Perfiles Educativos, 37*(147), 146-161. https://doi.org/10.22201/iisue.24486167e.2015.147.47271

Hernández-Infante, R. C., & Infante-Miranda, M. E. (2017). La clase en la educación superior, forma organizativa esencial en el proceso de enseñanza-aprendizaje. *Educación y Educadores, 20*(1), 27-40. https://doi.org/10.5294/edu.2017.20.1.2

Hernández Sampieri, R., Fernández Colado, C., & Baptista Lucio, M. P. (2006). *Metodología de la investigación*. McGraw Hill.

Hidalgo Troya, A. (2019). Técnicas estadísticas en el análisis cuantitativo de datos. *Revista Sigma, 15*(1), 28-44. https://revistas.udenar.edu.co/index.php/rsigma/article/view/4905

Higgins, S. (2013). Self regulation and learning: evidence from meta-analysis and from classrooms. In D. Whitebread, N. Mercer, C. Howe, & A. Tolmie (Eds.), *Self-regulation and dialogue in primary classrooms* (pp. 111-126). The British Psychological Society.

Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2015). *Conceptos básicos en evaluación como aprendizaje y empoderamiento en la educación superior*. EVALfor - Grupo de Investigación. https://doi.org/10.13140/RG.2.1.5070.5686

Ibarra-Sáiz, M. S., Lukas-Mujika, J. F., Ponce-González, N., & Rodríguez-Gómez, G. (2023). Percepción del profesorado universitario sobre la calidad de las tareas de evaluación de los resultados de aprendizaje. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa, 29*(1), 1-21. https://doi.org/10.30827/relieve.v29i1.27404

Iglesia Villasol, M. C. (2020). Evaluación formativa como elemento básico en la estrategia integral de aprendizaje ABPD, para futuros docentes. *Revista Iberoamericana de Evaluación Educativa, 13*(1), 71-92. https://doi.org/10.15366/riee2020.13.1.003

López-Pastor, V. M. (Coord.). (2006). *La evaluación en educación física: revisión de los modelos tradicionales y planteamiento de una alternativa: la evaluación formativa y compartida*. Miño y Dávila.

López-Pastor, V. M. (2007). La evaluación en educación física y su relación con la atención a la diversidad del alumnado. *Kronos: Revista Universitaria de la Actividad Física y el Deporte, 6*(11), 61-74.

López-Pastor, V. M. (2010) El papel de la evaluación formativa en la evaluación por competencias: aportaciones de la red de evaluación formativa y compartida en docencia universitaria. *REDU. Revista de Docencia Universitaria, 9*(1), 159-173. http://redaberta.usc.es/redu/index.php/REDU

López-Pastor, V. M., & Pérez-Pueyo, Á. (2017). *Evaluación formativa y compartida en educación: experiencias de éxito en todas las etapas educativas*. Ediciones Universidad de León.

López-Pastor, V. M., González Pascual, M. G., & Barba, J. J. (2005). La participación del alumnado en la evaluación: la autoevaluación, la coevaluación y la evaluación compartida. *Tándem Didáctica de la Edución Física, 17*, 21-37. https://dialnet.unirioja.es/servlet/articulo?codigo=1090624

López-Pastor, V. M., Hamodi Galán, C. M., & López Pastor, A. T. (2016). Evaluación formativa y compartida en educación superior: buenas prácticas y experiencias desarrolladas por la Red de Evaluación Formativa. In A. Pavié Nova, & M. Casas Uribe (Eds.), *Buenas prácticas de evaluación de aprendizajes en educación superior* (pp. 13-43). Universidad de los Lagos.

López-Pastor, V. M., Monjas, R., Manrique, J.-C., Barba, J. J., & González, M. (2008). Implicaciones de la evaluación en los enfoques de educación física cooperativa. El papel de la evaluación formativa y compartida en la necesaria búsqueda de coherencia. *Cultura y Educación, 20*(4), 457-477. https://doi.org/10.1174/113564008786542208

López-Rupérez, F. (2018). *Los porqués del MIR educativo*. Universidad Camilo José Cela.

López-Rupérez, F. (2021). La profesión docente en la perspectiva del siglo XXI: modelos de acceso a la profesión, desarrollo profesional e interacciones. *Revista de Educación, 393*, 69-96. https://recyt.fecyt.es/index.php/Redu/article/view/89652

Marshall, B., & Drummond, M. J. (2006). How teachers engage with assessment for learning: lessons from the classroom. *Research Papers in Education, 21*(2), 133-149. https://doi.org/10.1080/02671520600615638

Ocampo, J. D. F. (2017). La importancia de la evaluación para la mejora de la educación y así obtener calidad educativa. *Cuadernos de Educación y Desarrollo, 88*, 1-5.

Parra, S. G. (2013). SA y evaluación. Padres y Maestros. *Revista Padres y Maestros, 329*, 5-9. https://dialnet.unirioja.es/servlet/articulo?codigo=3144226

Poggi, M. (2008). Evaluación educativa. Sobre sentidos y prácticas. *Revista Iberoamericana de Evaluación Educativa, 1*(1), 37-44. https://doi.org/10.15366/riee2008.1.1.002

Prats, M. À. Sanmartí, N., & Oró, I. (2020). Evaluar para aprender con el apoyo de herramientas y recursos digitales. *Aloma: Revista de Psicologia, Ciències de L'educació i de L'esport Blanquerna, 38*(2), 9-20. https://doi.org/10.51698/aloma.2020.38.2.9-20

Real Decreto nº 157, de 1 de marzo de 2022. (2022, 1 de marzo). Por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. Ministerio de la Presidencia, Justicia y Relaciones con las Cortes. https://www.boe.es/eli/es/rd/2022/03/01/157

Richard, J. F., & Godbout, P. (2000). Formative assessment as an integralpart of the teaching-learning process. *Physical and Health EducationJournal, 66*(3), 4-9.

Rogero, J. (2015). Evaluar en la escuela pública liberadora e inclusiva. *Rescoldos: revista de diálogo social, 33*, 7-18.

Sanmartí Puig, N. (2019). *Evaluar y aprender: un único proceso*. Octaedro.

Santos Guerra, M. A. (1993). *La evaluación: un proceso de diálogo, comprensión y mejora*. Aljibe.

Santos-Guerra, M. A. (2003). *Una flecha en la diana: la evaluación como aprendizaje*. Narcea.

Sarduy Domínguez, Y. (2007). El análisis de información y las investigaciones cuantitativa y cualitativa. *Revista Cubana de Salud Pública, 33*(3), 1-11. https://doi.org/10.1590/S0864-34662007000300020

Seijo Galán, S., Freire Rodríguez, C., & Ferradás Canedo, M. M. (2020). Tipos de motivación en relación a la ansiedad ante los exámenes en el alumnado de educación primaria. *Publicaciones, 50*(1), 265-274. https://doi.org/10.30827/publicaciones.v50i1.14101

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record, 68*(7), 1-15. https://doi.org/10.1177/016146816706800707

Torrano-Martínez, R., Ortigosa-Quiles, J. M., Riquelme-Marín, A., & López-Pina, J. A. (2017). Evaluación de la ansiedad ante los exámenes en estudiantes de Educación Secundaria Obligatoria. *Revista de Psicología Clínica con Niños y Adolescentes, 4*(2), 103-110.

Torregrosa Martinez, A., Albarracín, L., & Deulofeu Piquet, J. (2021). Orientación y coevaluación: dos aspectos clave para la evolución del proceso de resolución de problemas. *Bolema: Mathematics Education Bulletin, 35*, 89-111. https://doi.org/10.1590/1980-4415v35n69a05

Villa Sanchez, A. (1985). La evaluación del profesor; perspectivas y resultados. *Revista de Educación, 277*, 55-93.

Wiliam, D., Lee, C, Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education, 11*(1), 49-65.

Zaragoza Costerad, J., Luis-Pascual, J. C., & Manrique Arribas, J. C. (2009). Experiencias de innovación en docencia universitaria: resultados de la aplicación de sistemas de evaluación formativa. *REDU. Revista de Docencia Universitaria, 7*(4), 1-33. https://doi.org/10.4995/redu.2009.6232

INFORMATIONS ABOUT THE AUTHORS

Noelia Santamaría-Cárdaba: Doctorate in Education from the University of Valladolid (Faculty of Education, Segovia). Graduate in Primary Education (UVa) and Sociology (UNED). Master's Degree in Educational Research (UVa). Currently a permanent lecturer in the Department of Education at the University of Valladolid. She has completed international stays in Bragança, Aveiro and Mexico, and nationally at the University of the Basque Country. She has published several articles in indexed databases such as JCR and SJR and several book chapters.
ORCID: https://orcid.org/0000-0001-6864-9330
E-mail: noelia.santamaria.cardaba@uva.es

Vanessa Ortega Quevedo: Doctor of Education from the University of Valladolid, specialising in science teaching, particularly in the nature of science and scientific thinking. She had an FPU contract at the University of Valladolid (Faculty of Education in Segovia) with a one-year postdoctoral fellowship. She is currently an Assistant Professor and works in the Department of Experimental, Social and Mathematical Science Teaching, specifically in the area of Experimental Science Teaching at the Faculty of Education-CFP at the Complutense University of Madrid. Her main lines of research focus on the development of scientific and critical thinking, the nature of science, STEAM education and climate change.
ORCID: https://orcid.org/0000-0002-5742-4678
E-mail: vanessao@ucm.es

Judith Cáceres Iglesias: Doctor of Education from the University of Valladolid (Faculty of Education, Segovia). She is currently an Assistant Professor in the Department of Educational Studies at the Complutense University of Madrid (Faculty of Education-CFP). She has worked on projects related to educational inclusion and has coordinated training courses on this topic. She has been a pre-doctoral FPU fellow at the University of Valladolid and a Juan de la Cierva researcher at the Autonomous University of Madrid.
ORCID: https://orcid.org/0000-0002-5831-6986
Correo electrónico de contacto: judcac01@ucm.es

NOTE:
The authors were responsible for designing, analysing and interpreting the data, as well as writing and critically reviewing the content of the manuscript and approving the final version.