

Assessment culture in highly effective schools in the Autonomous Community of the Basque Country

Izaskun Lopetegi-Mendizabal, Verónica Azpillaga-Larrea, Luis Lizasoain Hernández

Department of Education Sciences, Faculty of Education, Philosophy and Anthropology, University of the Basque Country, Donostia-San Sebastian, Spain

Article Info

Article history:

Received Jun 12, 2023

Revised Feb 4, 2026

Accepted Feb 23, 2026

Keywords:

Assessment culture
Assessment for learning
Evaluation
Formative assessment
Professional learning
School effectiveness
School improvement

ABSTRACT

The aim of the present study is to analyze the relationship between assessment culture and school effectiveness in schools identified as being highly effective in the Autonomous Community of the Basque Country (ACBC). In the first (quantitative) phase, 32 highly effective schools were selected using multilevel regression statistical procedures with hierarchical linear models. In the second (qualitative) phase, 90 semi-structured interviews were conducted with members of the management teams, inspectors and key consultants at the selected schools, focusing on a set of previously-established categories. In this paper, we analyze the categories linked to assessment. The results of the analysis suggest that highly effective schools use different formative assessment approaches to improve teaching and learning: data-driven decision making (DDDM), data-based decision making (DBDM), and assessment for learning (AfL). These approaches are developed differently in the two educational stages studied. A proposal is therefore made for helping them to move forward in this sense.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Izaskun Lopetegi-Mendizabal

Department of Education Sciences, Faculty of Education, Philosophy and Anthropology

University of the Basque Country (UPV/EHU)

Tolosa Hiribidea, 70 20018 Donostia-San Sebastian, Spain

Email: izaskun.lopetegi@ehu.eus

1. INTRODUCTION

During the 1990s, research in the field of education brought together two movements: school effectiveness and school improvement. The school effectiveness movement aimed to identify those elements that make schools more effective (i.e., capable of obtaining better student learning outcomes), whereas the school improvement movement focused on studying the processes that make schools better (in terms of student learning). Based on the idea that the two movements should work together to ensure complementary approaches, a theoretical-practical movement arose that drew on and received substantial contributions from both outlooks. This movement is known as effective school improvement (ESI). One pioneering work in this field was the improving school effectiveness project (ISEP), which was carried out between 1995 and 1999 in Scotland, under the coordination of MacBeath [1]. This was followed by a more ambitious project called capacity for change and adaptation of schools in the case of ESI [2], which involved eight research teams from different European countries. The result of this latter piece of research was the general framework for improving school effectiveness. The coordinators of the Spanish team were Muñoz-Repiso *et al.* [3] whose work in this field continues to bear fruit today [4]–[7].

The idea behind the efforts being made to improve school effectiveness is that in order to enhance student performance (effectiveness criterion), a series of intermediate aims must first be achieved (improvement criteria). To this end, schools that want to improve must first establish targets linked to

comprehensive student development and must designed and implement actions aimed at enhancing the quality of both the school and its teachers [4]. In any effort to enhance student outcomes, the key characteristics or effectiveness factors of the school and the classroom can be summed up, as in Table 1.

Table 1. The characteristics of effectiveness [4]

School	Effective classroom/teacher
Shared vision and aims	High expectations of students
Focus on outcomes and high expectations	Purposeful teaching
Professional leadership	Classroom climate
Teacher commitment and teamwork	Learning environment
Quality of the curriculum	Effective learning time/time dedicated to teaching and learning
Effective time management	Structured instruction
Professional teacher development	Independent learning
School climate	Differentiation and grouping procedures
Assessment potential	Progress monitoring
Engagement of families/school-family collaboration	Feedback and positive reinforcement
Good management of facilities and resources	

At the University of the Basque Country (UPV/EHU), we have been working in the field of ESI for over a decade. Our current project, entitled “school improvement from the gender perspective in schools in the Autonomous Community of the Basque Country (ACBC),” was preceded by another two projects entitled “characterization and best practices in high added value schools” and “longitudinal, contextualized study of schools with very high and very low levels of effectiveness. Design of school improvement actions”. All were carried out thanks to the collaboration of the Basque Institute for Research and Evaluation in Education (Basque Government) and funding granted by the Spanish Ministry for Science and Innovation.

In the first project, we identified several ‘best practice’ areas in which schools with high effectiveness levels in the ACBC stood out [8], [9]. The majority of the best practices identified have been disseminated through publications in national and international high-impact journals, and include: training and innovation, teacher training and professional development, methodologies, diversity, school organization and management, leadership, climate, and family-school-community relations. In this context, we also analyzed aspects linked to assessment, specifically assessment culture and its relationship with concrete students’ outcomes associated with effectiveness. One of the outcomes of this is the catalogue of best practices, analyzed by specific area: student, teacher and school. Best practices in the field of assessment in the student and school areas are as [8]:

“Special attention to student assessment, through frequent formative assessments with a clear, public and collegial definition of minimum contents and assessment criteria.”

“Assessment of all projects, programmers and activities engaged in by the school, with a positive evaluation and formative use of the diagnostic assessment.”

As these excerpts indicate, assessment is understood as a formative process that fosters student learning at both a classroom and school-wide level. School effectiveness is linked to a more developed assessment culture.

At the end of the 20th century, a deep-rooted change took place within the field of education. The first steps were taken towards shifting from a traditional framework focused on teaching and based on behavioral learning theory, to a contemporary one focused on learning and based on social constructivism and situated learning. A change also occurred in the construction of concepts such as teaching and learning; whereas in the traditional framework, high-quality teaching is understood as the transmission of specific knowledge, in the contemporary framework, it is conceived as the development of students who think. For its part, high-quality learning is understood in the traditional framework in terms of the acquisition of specific knowledge, whereas in the contemporary framework it is seen as the acquisition of problem-solving abilities [10], [11].

In this context, the term assessment culture refers to an approach to assessment that is closely linked to the type of teaching advocated by the contemporary paradigm of education [10], [12], [13] and is therefore different from testing culture, which is in line with the traditional paradigm. Testing culture views assessment as the last element in the teaching process. Its function is summative and its purpose is to assess learning by applying objective tests. Assessment culture, in contrast, views assessment as a systematic process that contributes to the teaching and learning process and is located at the point at which teaching and learning meet. Its function is formative and its purpose is to assess for learning. This type of assessment aims to ensure deep learning (developing the capacity for reflection and metacognitive strategies) and to enable feedback [11]–[13].

According to this approach, the purpose and function of assessment are distinguished as [14]: assessment for learning (AfL) is any assessment for which the top priority in its design and practice is to promote student learning and assessment really becomes formative when the evidence obtained is used to adapt the work to learning requirements. Formative assessment is characterized by its use of feedback strategies that provide teachers and students with information about their teaching and learning [15].

The term formative assessment encompasses a range of different approaches, some of which have very different conceptual bases [16]–[18]. These include: data-driven decision making (DDDM), data-based decision making (DBDM) and AfL; the last of which is further subdivided into three approaches: strategy-focused formative assessment, sociocognitive formative assessment and sociocultural formative assessment. DDDM is based on using interim assessments and state tests to improve teaching and learning. Teachers, individually or in teams, are expected to seek out and implement evidence-based instructional reforms. DBDM is based on using the data collected at the school to inform innovation decisions linked to teaching, curriculum and school performance. Strategy-focused formative assessment is based on using the strategies and tools of formative assessment to provide students with feedback and involve them more actively in their own learning. For its part, sociocognitive formative assessment is based on assessing students' understanding and skills as they engage in increasingly sophisticated practices common to disciplinary experts. Finally, sociocultural formative assessment is based on teachers helping students to reflect on how ways of knowing, doing and being in school relate to practices that are valued in their own families and communities.

Formative assessment focuses on three key questions: i) where are students heading; ii) where are they now; and iii) how can we close the gap between the two? [19]. Based on these questions, assessment that fosters student learning is made up of five key strategies [20]: i) clarify and share learning aims and criteria for success; ii) design effective classroom debates and other learning tasks to gather evidence of students' understanding; iii) provide feedback that helps students progress; iv) activate students as instruction resources; and v) activate students as the owners of their own learning. However, in order to promote learning through assessment practices, to integrating these practices into daily work, it is necessary to create a suitable environment that encourages learning: developing norms of mutual respect, personal responsibility and shared focus on learning, establishing a healthy relationship between formative and summative assessment and avoiding grading practices [21]. At the end of the day, assessment culture is defined in terms of three interrelated components [22]: i) teachers' and students' beliefs about the aims of the assessment and its relationship with teaching and learning; ii) the assessment practices used by teachers, in which students are involved; and iii) the assessment tools that support, or in some cases, inhibit these practices.

Promoting an assessment culture requires a reculturation of assessment practices that have been embedded for a long time in the practices of teachers and schools. This reculturation occurs over time and requires effective evidence [23]. To achieve this, the focus of learning must be on teacher training; this is what is advocated by the researchers who have developed the different models currently driving the incorporation of assessment culture [11], [15], [19], [24], [25]. Those who have been engaged in this undertaking in schools over recent years remark that, for this reculturation to occur, a series of conditions must be met [15], [23], [24], [26], [27]. Teachers must learn new AfL practices and understand that the purpose of assessment is learning; they must reflect on practices in collaboration with other teachers as they incorporate AfL into their daily work; and they must be aided in this task by experts.

This work is linked to the learning process of professional learning groups, which are characterized by their focus on collaboration, their shared perspectives and aspirations, their focus on student learning and their use of reflective dialogue in teacher learning [28]–[30]. The effectiveness of this practice in developing the AfL in the classroom has recently been demonstrated [31], [32]. Other studies have identified several essential prerequisites for this reculturation to occur, including goal setting, pedagogical content knowledge, assessment literacy, feedback strategies, the facilitation of classroom discussions, involving students, collaboration between teachers, and a positive attitude around AfL [18].

Teachers develop the ability to use AfL in a teacher professional learning continuum, until they achieve the assessment culture mindset. Five levels have been identified in this progression: i) learning the letter: when teachers begin to connect learning with assessment (novice); ii) practicing the letter: when teachers implement AfL practices and start asking questions about the procedure; iii) responding to the letter: when teachers are finding strengths and gaps in their knowledge and practice of AfL, and are looking for answers to them; iv) adopting the spirit: when teachers come to truly believe that the relationship between instruction, AfL strategy, practice and assessment culture fosters student learning; and v) leading the spirit: when teachers help other teachers to engage in AfL by modelling the AfL spirit in their classrooms [26].

Developing an assessment culture in schools is no easy task, since four components of a complex system interrelate with each other in a process of co-evolution [33], [34]: mindset, learning in the classroom, professional teacher learning, and leadership. The characteristics of a developed assessment culture are as: learning-centered paradigm; assessment guides teaching and learning; assessment means interaction/dialogue

with students and involves empowering students; diversity is desirable and assessment is carried out regularly, in an informal manner, through dialogue; individual and collective self-efficacy, understood as belief in individual and collective capabilities, are promoted; and tests are not necessarily considered a trustworthy indicator [33], [34]. In a highly developed assessment culture, classroom learning is characterized by optimal AfL, which corresponds to a research cycle [35] and is fostered in a learning environment that offers support mechanisms and cultivates rules, values and beliefs that are conducive to learning [35], [36]. In a highly developed assessment culture, professional teacher learning is characterized by a focus on student learning, a shared school vision, reflexive dialogues, collaboration, shared responsibility coupled with high expectations regarding the learning of all students at the school, professional self-efficacy and collective efficacy, and a favorable social climate [33]. In addition, leadership is characterized by the fact that the head teacher is seen as a pedagogic leader whose priority is the professional development of individual and collective capabilities, and who practices distributed leadership supported by existing rules and the acceptance of mistakes [10].

During the period of analysis contemplated in the present study, assessment in schools located in the ACBC was governed by Organic Law 2/2006, of 3 May, on Education (LOE) [37] and Decree 175/2007, of 16 October, establishing the curriculum for basic education in the ACBC [38]. This law also stipulates that a diagnostic assessment (DA) of students' basic competences shall be carried out at the end of the second two years of primary and in year 2 of compulsory secondary education. This assessment is formative in nature and serves to guide schools and inform families and the entire educational community (Art.36). The aim of the DA is to foster the continuous improvement of teaching-learning processes and the education system [39].

The aim of the present study is to analyze the relationship between assessment culture and school effectiveness in schools identified as being highly effective in the ACBC by searching for the characteristics of a developed assessment culture. The research questions central to this study are:

- What characteristics of formative assessment are found in highly effective schools?
- How do the characteristics of formative assessment relate to the components of a developed assessment culture?
- What characteristics of a developed assessment culture are found in highly effective schools?

2. METHOD

An explanatory, sequential, mixed methods design was used in the present study [40]. During the first (quantitative) phase, a correlational study was conducted using multilevel regression statistical procedures performed through hierarchical linear models; and in the second (qualitative) phase, a multiple case study design was followed.

2.1. Quantitative phase

To detect and select highly effective schools, we used hierarchical linear models to statistically control for the effect of contextual covariables. This enabled us to isolate the effect of the school itself and to consider it an indicator of effectiveness. This type of regression model simultaneously takes into consideration and studies the effect of multiple variables operating at different levels. In this study we consider two levels: student level and school level.

Participants in the quantitative phase were all the schools located in the ACBC, since the DA is a census-based assessment carried out in the 4th year of primary education and the 2nd year of compulsory secondary education. Specifically, 409 primary schools (PS) and 324 secondary schools (SS) (providing compulsory secondary education) were included in the study. The DA data collected in 2009, 2010, and 2011 in the basic skills of mathematics and linguistic communication in Spanish and Basque were analyzed. A regression model was carried out for each competence, year group and year evaluated; that is, three competences, 2-year groups and 3 years, resulting in a total of 18 regression models.

The criterion variable used was the scores obtained by students in the DA mathematics and Spanish and Basque language communication tests. The contextual covariables were extracted from the information provided in the context questionnaires administered alongside the DA tests by the Basque Institute for Research and Evaluation in Education (ISEI/IVEI). The covariables analyzed were:

- At the student level (L1): language model in which the student was studying (A, B, D), gender, whether or not Basque was habitually spoken at home, whether or not the student was an immigrant who had entered the education system later than other students, the family's economic, social and cultural status (ESCS), previous performance (grade achieved the previous year in the subject, as reported by students themselves) and whether or not the student had repeated an academic year (normative age).

- At the school level (L2): type of school (public or semi-private) and aggregated L1 variables (mean ESCS of the school, percentage of immigrant students at the school, rate of Basque use at home, retake rate, and mean previous performance).

The selection criterion used was schools with a high positive residual [8]. Residuals were calculated using the equation at level 2 of the multilevel regression model, in which the effectiveness criterion (u_{0j}) was the residual of each school associated with the intersection (β_{0j}), after having controlled for the individual and school factors included in the model. The general notation for this model [9], at the student level (N1):

$$Y_{ij} = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij} + r_{ij}$$

At the center level (N2):

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \sum_{s=1}^S \gamma_{0s} W_{sj} + u_{0j} \\ \beta_{qj} &= \gamma_{q0} + \sum_{s=1}^S \gamma_{qs} W_{sj} + u_{qj} \end{aligned}$$

Where, β_{0j} : mean performance of school j , γ_{00} : effect common to all schools, γ_{0s} : linear effect of covariable W_s on mean school performance, W_{sj} : that value taken by school j in the school covariable W_s , u_{0j} : residual for each school. As the central aim of the paper is not the statistical modelling itself, these models are described briefly, and readers are referred to other works [8], [9].

2.2. Qualitative phase

In order to identify and describe the practices associated with schools deemed to be highly effective, we used a multiple case study design [41] with a qualitative approach. Participants were 90 informants (internal and external) from the schools selected during the quantitative phase and included members of the management teams and school inspectors and consultants assigned to the schools in question. Data was gathered during a series of semi-structured interviews focusing on a set of previously established categories: i) projects, plans and training; ii) teaching methodologies and materials; iii) attention to diversity; iv) student monitoring, individual attention, guidance and tutorials; v) student assessment; vi) time management; vii) leadership and the management team; viii) management and organization models; ix) coordination; x) engagement and sense of belonging; xi) teacher appraisal, school evaluation, evaluation of programmers and activities and use of the DA; xii) school climate; xiii) image, facilities and resources; and xiv) family, community and social capital. Interviews were audio recorded and the information selectively transcribed in the assigned categories and subcategories.

The qualitative analysis presented was carried out with the fifth and eleventh categories: ‘student assessment’ and ‘teacher appraisal, school evaluation, evaluation of programmers and activities and use of the DA’. A qualitative mixed categorical analysis of the content of these two areas was conducted using inductive and deductive categories [42]. For the inductive categorization, we combined the information pertaining to the two categories and analyzed the discourse of the professionals interviewed. The following coding system was used to differentiate between the educational stages studied and the type of professionals interviewed: PS, SS, member of the management team (MAN), inspector (INS), and consultant (CON). After the inductive analysis of the information, we carried out the deductive categorical analysis with the aim of identifying the formative assessment processes that led to the development of an assessment culture. To this end, we used the categories scheme proposed by Birenbaum [33], [34], which is shown in Table 2.

Table 2. Category scheme for a developed assessment culture

Sub-categories	Components
Mindset	Learning-centered paradigm Assessment guides teaching and learning Assessment means dialogue/interaction with students Assessment empowers students Diversity is desirable Individual and collective self-efficacy Modesty is required in assessment
Learning in the classroom	Assessment for optimal learning Constructivist classroom culture
Professional teacher learning	Professional collaboration for learning Positive social climate at work Contextualization
Leadership	Facilitator of school-based professional learning Organizational structure conducive to distributed leadership Open attitude to inquiry and ongoing reflection

3. RESULTS AND DISCUSSION

3.1. Results of the quantitative phase

The intraclass correlation coefficient (ICC) index values, coefficients of each covariable and the explanatory capacity of the model changed in each model analyzed. No great difference was observed between the different years analyzed, although differences were found between primary and secondary and, above all, between subjects. The mean ICC was 16%. The analysis of the results revealed that the most important covariables at level 1 (students) were ESCS, being an immigrant, language spoken at home and retaking a year; at level 2 (schools), the most important covariable was mean school ESCS. In relation to linguistic communication in Basque, the language model in which students studied was particularly important. In terms of the explanatory capacity of the models, it should be noted that although they have a limited explanatory capacity in relation to the variance observed between students (around 12% at the most), this figure is much higher in relation to the variance observed between schools (level 2), being as high as 75-80% in this case. For further information, readers can consult works [43].

Based on the results obtained following the statistical analysis, we selected only those schools in which a high mean residual (over the 80th percentile) had been obtained for both academic years in all three basic competencies or in two of them and in the mean for all four subjects measured. A total of 32 schools were selected. The selected schools were evenly distributed within the total population of schools in the ACBC, as shown in Table 3. A total of 16 were PS (8 public and 8 semi-private) and 16 were SS (8 public and 8 semi-private). Their specific characteristics were as: 7 had a low ESCS (3/4), 17 a medium ESCS (9/8) and 8 a high ESCS (4/4); and 19 had a low percentage of immigrant students (10/9), 5 a medium percentage (2/3) and 8 a high percentage (4/4).

Table 3. Characteristics of highly effective schools

PS (16)		N		%		Compulsory SS (16)	
Characteristics		N	%	Characteristics		N	%
Type of school	Public	8	50	Type of school	Public	8	50
	Semi-private	8	50		Semi-private	8	50
School's socioeconomic and cultural status	Low	3	18.75	School's socioeconomic and cultural status	Low	4	25
	Medium	9	56.25		Medium	8	50
	High	4	25		High	4	25
Percentage of immigrant students	Low	10	62.5	Percentage of immigrant students	Low	9	56.25
	Medium	2	12.5		Medium	3	18.75
	High	4	25		High	4	25

3.2. Results of the qualitative phase

3.2.1. Emergent analysis

We will now present the results of the emergent analysis of the information. Figure 1 shows the categorization of the management team, consultant and inspector discourses. At first glance, it is clear that the discourse of management teams focused mainly on two key categories: assessment process and improvement. The first encompassed all the descriptions given of the student assessment process, and the second all the information provided regarding efforts to improve teaching. Consultants talked mainly about approaches linked to the DA, under the category of the same name and inspectors talked mainly about evaluation in more general terms, again under the category of the same name.

Within the assessment process category, the subcategory assessment criteria was of particular importance at both primary (8) and secondary level (11). This subcategory encompassed all efforts aimed at identifying and clarifying assessment criteria. The next most frequently mentioned subcategory was individualization in primary (7) and individualization/adaptation in secondary (6). Another subcategory that is worth mentioning due to the number of references made to it at the primary level is coordination (6).

In the improvement category, information was provided on efforts aimed at improving teaching. Within this category, the subcategories use of the DA at primary level (9) and internal evaluation at secondary level (13) were of particular importance. References to these two subcategories at the other level, i.e., use of the DA at secondary level (3) and internal evaluation at primary level (4), are also worth noting. The subcategory use of the DA outlined how schools used the DA to promote learning.

The subcategory internal evaluation included descriptions of the internal evaluation processes established at the schools. The analysis of management team discourses revealed that internal evaluation processes were in place in most highly effective SS (13/16). Although no two processes were exactly alike, in general, they can be divided into three main groups: some schools (3) have their own systematic evaluation process that has been adapted to their specific educational project, others (5) use the processes outlined in the quality system, regardless of whether or not they are currently in said programmed; and others (5) focus on

assessment and propose improvements separately within each department. Consultant discourses focused on how processes linked to the diagnostic assessment were carried out in the highly effective schools analyzed. Inspector discourses focused on the importance attached to evaluation at the schools studied and the way in which it is used.

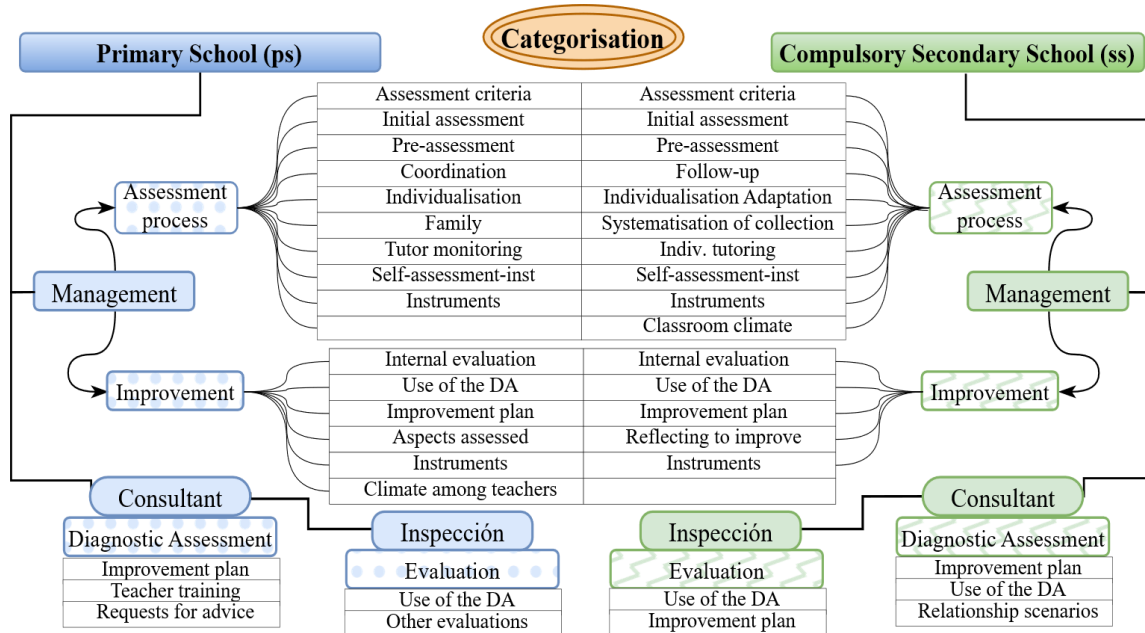


Figure 1. Discourse category

3.2.2. Deductive analysis

To carry out the deductive analysis, we reordered the subcategories in accordance with the categories proposed by Birenbaum [33], [34] and identified indications of the characteristics linked to a developed culture of assessment. Table 4 shows the characteristics of a developed assessment culture [33]–[36] in relation to three components, along with the subcategories of the emergent analysis in which said characteristics appeared. The components identified were learning in the classroom, professional teacher learning and leadership.

Table 4. Assessment culture and emergent categories

Components	Characteristics	Primary	Secondary	
Learning in the classroom	Assessment for optimal learning	Planning	<Assessment criteria>	<Assessment criteria>
		Instrument design	<Instruments>	<Instruments>
		Evidence collection and interpretation	<Initial assessment, pre-assessment and coordination>	<Initial assessment and pre-assessment>
		Use	<Individualization, tutor monitoring and family>	< Individual., Adaptat., Monitoring and indiv. tutoring>
		Evaluation	<Internal evaluation and reflection>	
Professional teacher learning	Constructivist classroom culture Professional collaboration for learning	Individual level	<Self-assessment-inst>	<Self-assessment-inst>
			-----	<Classroom climate>
			<Internal evaluation and Reflection>	<Reflecting to improve>
			<Use of the DA>	<Use of the DA>
			<Improvement plan>	<Improvement plan>
Leadership	Positive social climate at work Contextualization Pedagogical leader Distributed leadership Open attitude to inquiry and ongoing reflection		<Climate among teachers>	-----
			-----	-----
			-----	-----
			<Internal evaluation and reflection>	-----

a. Learning in the classroom

The learning in the classroom component encompasses two characteristics of a developed assessment culture: assessment for optimal learning and constructivist classroom culture, as in Table 4.

– Assessment for optimal learning

When we talk about AfL in the classroom, it should be noted that we are in fact talking about two separate processes. The assessment process carried out by the teacher and the assessment process carried out individually by each student. Assessment for optimal learning includes both of these processes. An assessment for optimal learning cycle is a research cycle comprising five phases: planning, designing instruments to measure student understanding, evidence collection and interpretation, use and evaluation. At an individual level, students use a similar process to regulate the learning process [35], [36].

Planning refers to the process of establishing and defining learning goals and targets. It is associated with the tasks aimed at identifying and clarifying assessment goals outlined in the subcategory assessment criteria. This practice was identified in the initial work carried out by the research team as a more generalized best practice linked to student assessment. Three aspects were outlined [8]: the minimum level required should be clearly defined and should be agreed upon in a collegiate manner by all members of the teaching staff or by a team of teachers, and communicated to the different interested parties (families and students) through diverse channels (noticeboards, correspondence, and website).

Although some mention was made by PS informants of the importance of students being aware of the assessment criteria, in general, at this level, much more value is attached to the work carried out by the teaching staff as a whole to establish and reach a consensus regarding assessment criteria and the minimum level required, since this aspect was mentioned in almost all interviews held. In contrast, at secondary level, much more emphasis was placed on the importance of all students being aware of the assessment criteria.

“Students know, right from the start, what is being assessed. That established in the different departments is later agreed upon at each level; it is discussed by the whole team.” (5ps-MAN)

“An agreement is reached among all teachers regarding assessment criteria.” (10ps-MAN)

“Teachers know what is going to be assessed.” (1ps-MAN)

“We are working to ensure that all students are aware of the assessment criteria, right from day one, which is why teachers publish them and explain them to their students. The criteria are posted on the noticeboard in each classroom, as well as on the school’s website.” (13ss-MAN)

Several tools were observed that aimed to measure students’ progress in their learning pathway. Primary level informants did talk about tests and exams (12ps-MAN), minimum level records and tests (13ps-MAN; 6ps-MAN), observation scales (7ps-MAN) and observation records for monitoring purposes (10ps-MAN; 6ps-MAN); and secondary level informants referred to workbook monitoring (2ss-MAN) and exam tests (2ss-MAN; 14ss-MAN).

Informants mentioned several moments in the academic year at which they gather or interpret evidence in order to adapt the teaching-learning process. Primary level management teams said they collected evidence of students’ current location on their learning pathway at the start of each academic year. They then monitor their progress using observation records kept by teachers and meet to check how each student is doing in terms of assessment and to analyze which teaching strategies are working in each case.

“We carry out the initial assessment at the beginning of the academic year, and then once every term after that. The observation records kept by each teacher are the basis for all student monitoring. We have days set aside for meetings aimed at determining how each student is doing in terms of assessment.” (10ps-MAN)

At the secondary level, informants claimed that assessment was systematic and carried out on a continuous basis. They also said that they gathered evidence regarding each student’s progress (although they did not specify how) and that this evidence was used to adapt teaching programmes. Alternatively, some schools carry out a diagnosis of needs at the start of the academic year, which is then used to organize the direct support required in each specific area. In some schools, informants stated that the needs analysis is only carried out when a student is observed not to have reached the established minimum competence level.

“In terms of assessment, it’s all written down. Assessment is ongoing and is determined systematically.” (8ss-MAN)

The SS management teams also talked about pre-assessments, i.e., the gathering of data prior to the assessment that serves as a guide for teachers, indicating those areas in which students need further instruction before the learning assessment itself:

Assessment culture in highly effective schools in the Autonomous ... (Izaskun Lopetegi-Mendizabal)

“Every term we carry out a pre-assessment, followed by the assessment itself.” (9ss-MAN)

However, the management teams said they did not know whether or not feedback was provided after the pre-assessments, commenting that one would have to consult the teachers themselves.

Several feedback strategies were identified that serve to help and guide students along their learning pathway or, as other authors would say, to close the gap between expected and real results. Primary level management teams said that they carried out individual assessments - an initial one at the beginning of the academic year and then a follow-up to monitor progress; they also stated that each student and their family were told where they were on their pathway and where they should be heading. They also said they used observation records to monitor students:

“The observation records kept by each teacher are the basis for all student monitoring.”
(10ps-MAN)

Also, that teacher collaboration was important to improve teaching:

“Assessments are carried out individually with each child. Meetings are arranged for this purpose, in which all the teachers working with each class analyze what is working in each case.” (8ps-MAN)

Also, establish guidelines for collaborating with families to foster learning at home:

“We give families guidelines to help them work at home.” (4ps-MAN)

Informants also talked about continuous formative assessment:

“We carry out an assessment once a fortnight, analyzing the result and designing an intervention, with the corresponding feedback.” (1ps-MAN)

Secondary level management teams (15ps-MAN) talked about monitoring through the collection of informal evidence, *“close monitoring, being alert,”* and making changes to the teaching programmers, *“we end up adapting to the kids”* (9ps-MAN). In terms of feedback to students, informants from one school talked about systematic feedback through individual tutoring sessions in which participants highlight each student’s achievements and set goals for the future:

“Individual tutoring is more an exercise in valuing what has been achieved than assessing progress. You positively value everything the student has achieved and establish the next set of goals they need to work towards.” (1ss-MAN)

Evaluation refers to the process of ascertaining the effectiveness of the measures taken to close the gap between expected and real results. A good example was given by the management team of a PS, who described a system in which each teacher was encouraged to think about their actions with each student (presented in coordination meetings) and their praxis in more general terms (presented in the form of best practices at the end of each academic year).

“Evaluations are carried out individually with each child. Meetings are arranged for this purpose, in which all the teachers working with each class analyze what is working in each case.” (8ps-MAN)

“At the end of each academic year, each teacher has the opportunity to highlight something they feel they have done well, or in other words, their “best practices”.” (3ps-MAN)

AfL fosters the capacity for self-assessment with the aim of developing reflection and self-management skills, as well as encouraging students to become instruction resources for each other, so they can progress in their own learning.

“Student self-assessment in all activities.” (2ps-MAN)

“Self-assessment to foster autonomy, self-demandingness, self-regulation and the recognition of mistakes.” (5ps-MAN)

“Students engage in self-assessment.” (12ss-MAN)

“Before [...] you could never get students to share with each other [...] but then, they gradually realized that by explaining and sharing, they actually learned more.” (1ss-MAN)

– Constructivist classroom culture

This is a type of classroom culture that fosters student learning and reflects constructivist ideas regarding instruction, learning and assessment. The characteristics of this kind of culture include those linked to a climate of collaboration. We observed beliefs conducive to this kind of culture in one of the SS analyzed.

“We didn’t have a bad climate, but now we see that the relationships have improved. Before, you managed to get everyone to respect each other in the classroom, but you could never get students to share with each other. It is normally those students who have a higher level who find it hardest to share. But then, they gradually realized that by explaining and sharing, they actually learned more.” (1ss-MAN)

b. Professional teacher learning

The professional teacher learning component encompasses three characteristics of a developed assessment culture: professional collaboration for learning, positive social climate at work and contextualization, as in Table 4.

– Professional collaboration for learning

The assessment process for teacher learning is also a research cycle. Teachers work together to explore their own praxis. Evidence of collaborative research to improve one’s own praxis was found among informants working in primary education. In SS, collaborative reflection among teachers with a view to improving praxis is mainly carried out at a departmental level.

“We have even recorded classes to analyze certain interactive aspects, with a view to improving teaching efficacy.” (5ps-MAN)

“Meetings are arranged for this purpose, in which all the teachers working with each class analyze what is working in each case.” (8ps-MAN)

“In addition to coordinating efforts and analyzing what is happening, they also ask themselves what they can do to improve.” (4ss-MAN)

Another means of fostering the ongoing improvement of teaching-learning processes in the ACBC is the DA [39]. Evidence of the formative use of this assessment process was found in both primary and SS. Most of the management teams working at the primary level viewed the DA in a positive light and said they used it to introduce improvements. They claimed it was an *“opportunity, offering guidance and information for improvement”* (7ps-MAN). They also said that *“thanks to the DA they have started to adopt a more competence-based approach”* (16ps-MAN) and that *“the DA prompted them to stop and think about how to do things and how to assess them”* (3ps-MAN). Furthermore, the SS management teams said that they used the DA to propose improvements linked to the strategic areas. External informants described how these improvements were implemented.

“They analyze the results of the DA in detail. They analyze each point, all the characteristics, and each specific student.” (1ss-INS)

“They are interested in hearing about how they can improve. They are given the DA report so that can analyze it and they are told about the improvements they could make or implement. They know how to make the most of the advice provided by the Berritzegune. They use strategies to make the most of it.” (1ss-CON)

– Positive social climate at work

In order for teachers to engage in a collaborative analysis of their own praxis, there needs to be a social climate at the school that is conducive to this sort of process. The characteristics of a positive social climate at work include transparency and privacy rules, and the acceptance of mistakes. Informants from the Berritzegune consultancy service working at the primary level made several comments linked to this climate.

“The core teaching team, the members of which have been at the school for years, play a key role in transmitting the pedagogical values. [...] They believe in what they do; and teacher commitment, systematization and task coordination are considered to be of the utmost importance. They are very participatory and engage in innovative projects; the climate is very good and the management offers leadership.” (10ps-BER)

– Contextualization

Contextualization refers to basing professional learning on students' specific learning needs, and involves complete research cycles. This component was not found in the information analyzed.

c. Leadership

The leadership component includes three characteristics of a developed assessment culture: leadership that facilitates school-based professional learning (pedagogical leadership), organizational structure conducive to distributed leadership, and open attitude to inquiry and ongoing reflection, as in Table 4. In the information analyzed, indications of an open attitude to evaluation and constant reflection were found at the primary level. No indications of pedagogical leadership or distributed leadership were found.

“We assess everything: overall school results, subject results, differences between teachers. And this has prompted us to engage in ongoing reflection that enables teacher training.” (5ps-MAN)

d. Mindset

AfL at school results in the development of a specific teacher mindset (fourth component) that is characterized by some assessment attributes. Certain indications of these attributes were found in the analysis carried out, as in Table 5: assessment guides teaching and learning in PS, assessment guides teaching in SS, assessment empowers students in PS and SS, and individual and collective self-efficacy are fostered in SS.

Table 5. Attributes of the assessment culture mindset found

Attributes	Expressions and practices	
	Primary education	Compulsory secondary education
Assessment is about learning		
Assessment guides teaching and learning	<i>“Meetings are arranged for this purpose, in which all the teachers working with each class analyze what is working in each case.”</i> (8ps-MAN) <i>“Pre-assessments, assessments, child by child, monitoring, seeing how they progress, whether or not something needs to be done.”</i> (7ps-MAN)	<i>“We have some minimum competence levels for the different subjects, but at the end of the day, we end up adapting to the kids.”</i> (9ss-MAN) <i>“The kids set the pace.”</i> (6ss-BER)
Assessment means dialogue/interaction with students		
Assessment empowers students	<i>“Student self-assessment in all activities.”</i> (2ps-MAN) <i>“They try to use self-assessment and co-assessment.”</i> (6ps-MAN) <i>“We attach a great deal of importance to assessment. We base everything we do more on values than on results. Not only in the external assessment, but in the internal one also (autonomy, being self-demanding, students’ capacity for self-regulation, and identifying mistakes).”</i> (5ps-MAN)	<i>“Students engage in self-assessment.”</i> (12ss-MAN)
Diversity is desirable		
Individual and collective self-efficacy		<i>“Another value is being demanding and disciplined, and developing good work and study habits from an early age. Being demanding means you have to try to be the best you can possibly be. We have high expectations of our students.”</i> (3ss-MAN)
Modesty is required in assessment		

3.3. Discussion

Following the analysis of the improvement category, and taking into account the different approaches under the umbrella term formative assessment [16]–[18], we can conclude that school improvement in the highly effective schools included in our study is guided by formative assessment processes that follow the DDDM and DBDM approaches. In primary education, DDDM processes are the most common, with the information from the DA being used to promote improvements in teaching and learning. In secondary education, both types of formative assessment processes are used, although the tendency leans more towards DBDM, which prioritizes the use of data collected within the school to inform any improvement decisions to be made.

The analysis carried out to identify instances of the AfL approach being adopted in highly effective schools revealed that, in these schools, an effort is being made to implement AfL with a formative function in

the classroom. However, given the use made of formative assessment at the school level in the schools analyzed, few AfL practices were identified in the classroom and the assessment culture mindset had only been developed to a limited extent. This seems to suggest that highly effective schools are in the early stages of developing an assessment culture, or in other words, are still at the learning the letter or practicing the letter stage [26]. There seems to be something hindering the development of an assessment culture in schools, as indeed has been suggested by certain studies that report that teachers face different challenges when implementing AfL in their classrooms. For example, Gebremariam and Gedamu [44] found that although teachers used AfL practices in primary education classrooms to a high or moderate degree, they nevertheless faced general challenges linked to transparency, experience, training, school problems, and trust in preferences. For his part, Kanjee [45] found that although teachers did indeed employ AfL practices in their classrooms, very few were able to demonstrate their effective use. The number of currently coexisting models that aim to promote the development of an assessment culture in the classroom may also be an indicator of this added difficulty [11], [15], [20], [24].

Over recent years, research has provided some specific examples of contextualized AfL practices that may help teachers implement AfL in their classroom. In primary education, for example, Fives and Barnes [46] identified a series AfL implementation, interpretation and return routines from a fifth grade classroom, and Mohamed *et al.* [47] pinpointed four key strategies characterizing high-quality AfL practices in English classes: shared learning objectives, use of different types of questions to assess student understanding, providing constructive feedback to even the best student in the class and offering guidance to students to help them produce their own progress reports. In secondary education, Adie *et al.* [23] found that developing AfL practices at school requires a major effort in terms of complex cultural leadership, since it involves implementing new practices with teachers, students and families.

Previous studies have explored the factors required for implementing AfL in the classroom [18], finding that teacher collaboration is a key prerequisite. The results of our analysis revealed that highly effective schools have teacher collaboration teams aimed at improving teaching and learning in both primary and secondary classrooms. In primary education, the teams were found to engage in reflective practices through collaborative learning, with the aim of modifying the teaching and learning process, and some groups even engaged in the practice of lesson study. In secondary education, the different departments were found to function as teacher collaboration teams with the aim of promoting the improvement of teaching and learning. Reflective practice is a professional learning process that facilitates the pedagogical development of teachers in action through various strategies, including collaborative learning among teachers [48]. Existing research supports the effectiveness of this type of collaborative reflective practice in improving classroom instruction [49]. Lesson study is a collaborative learning inquiry cycle practice that involves implementing lessons in the classroom that enable teachers to collectively achieve specific learning objectives [50], [51]. This collaborative learning technique has a significant influence on teachers' self-efficacy [52].

4. CONCLUSION

The aim of the present study was to analyze the relationship between assessment culture and school effectiveness in schools identified as being highly effective in the ACBC. The search for characteristics linked to the components of a developed assessment culture has revealed that, in highly effective schools, work is being done to implement the formative assessment approach called AfL and that said assessment has different characteristics depending on the stage to which it belongs. At the classroom level, the most important distinctive characteristic is that, in primary education, the information gathered is used to provide students with feedback and to improve teaching, whereas at the secondary level, it is mainly used to improve teaching. At the professional learning level, in the PS studied, we found indications of teacher research and collaboration aimed at improving student learning; whereas in the SS studied, we found indications of teacher collaboration for improving praxis, which is carried out at a departmental level. And at the mindset level, evidence was found in primary education of a belief that assessment guides teaching and learning, and in SS, evidence was found of a belief that assessment guides teaching. Nothing in the interviews indicated the presence of the belief that assessment is about learning, or that it means dialogue/interaction with students.

In light of the development of certain components of assessment culture in highly effective schools located in the ACBC, it seems that PS are closer to said culture than schools working at the level of compulsory secondary education. Consequently, bearing in mind the fact that the reculturation of assessment practices is a difficult process, and following the advice given by specialists in the field, we propose that, in order to further develop assessment culture in schools, an AfL inquiry-based professional learning processes be set in motion among the assessment teams that are already in place in the highly effective schools studied.

ACKNOWLEDGEMENTS

The present study was made possible by the projects ‘Characterization and best practices in high added value schools’ [EDU2011-24366] and ‘School improvement from the gender perspective in schools in the Autonomous Community of the Basque Country’ [PGC2018-094124-B-100, MCIU/AEI/FEDER, UE], and the research group of UPV/EHU GANDERE [GIU21/056]. This line of work is carried out with the collaboration of the Basque Institute for Research and Evaluation in Education (ISEI-IVEI).

FUNDING INFORMATION

This work was supported by Ministerio de Economía y Competitividad (Spain) under the National R&D Plan (EDU2011-24366) and by the Spanish Ministry of Science, Innovation and Universities (MCIU), the State Agency for Research (AEI) and the European Regional Development Fund (ERDF) through the project PGC2018-094124-B-I00.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Izaskun Lopetegi-Mendizabal	✓	✓			✓	✓		✓	✓	✓	✓			
Verónica Azpillaga-Larrea		✓				✓	✓	✓		✓		✓		✓
Luis Lizasoain Hernández		✓				✓	✓	✓		✓			✓	✓

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : **O**riginal Draft

E : **E**diting

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors’ institutional review board or equivalent committee.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.




REFERENCES

- [1] J. MacBeath, "Improving school effectiveness: retrospective and prospective," in *International Handbook of School Effectiveness and Improvement: Review, Reflection and Reframing*, T. Townsend, Ed. Dordrecht: Springer Netherlands, 2007, pp. 57–74, doi: 10.1007/978-1-4020-5747-2_4.
- [2] B. P. M. Creemers, L. Stoll, and G. Reezigt, "Effective school improvement—ingredients for success: the results of an international comparative study of best practice case studies," in *International Handbook of School Effectiveness and Improvement: Review, Reflection and Reframing*, T. Townsend, Ed. Dordrecht: Springer Netherlands, 2007, pp. 825–838, doi: 10.1007/978-1-4020-5747-2_44.
- [3] M. Muñoz-Repiso *et al.*, "Improving processes, improving educational outcomes European research on improving school effectiveness. Madrid, España: Ministerio de Educación, Cultura y Deporte, ICE de la Universidad de Deusto y Ediciones Mensajero (in Spanish), 2003. [Online]. Available: <https://dialnet.unirioja.es/servlet/libro?codigo=325136>
- [4] F. J. Murillo, "Improvement of the school effectiveness in Latin America," (in Spanish), *Revista Iberoamericana de Educación*, vol. 55, pp. 49–83, Jan. 2011, doi: 10.35362/rie550525.
- [5] F. J. Murillo and C. Martínez-Garrido, "Factors of academic efficacy in the Dominican Republic," (in Spanish), *Innovación Educativa*, vol. 16, no. 72, pp. 113–132, 2016.
- [6] C. Martínez-Garrido and F. J. Murillo, "Research on effective teaching. A multilevel study for Ibero-America," *Educación*, vol. 31, no. 61, pp. 46–75, Oct. 2022, doi: 10.18800/educacion.202202.003.
- [7] F. J. Murillo and N. Hidalgo, "Fair student assessment: A phenomenographic study on teachers' conceptions," *Studies in Educational Evaluation*, vol. 65, p. 100860, Jun. 2020, doi: 10.1016/j.stueduc.2020.100860.
- [8] L. Lizasoain *et al.*, "School effectiveness in schools in the Basque Country. Final Report (2011–2015)," (in Spanish), 2015. [Online]. Available: https://isei-ivei.euskadi.eus/documents/d/guest/la-eficacia-escolar-en-los-centros-del-pais-vasco_definitivo?imagePreview=1
- [9] L. Lizasoain *et al.*, "Report on school effectiveness and improvement in the Basque Autonomous Community (2015–2019)," (in Spanish), 2022. [Online]. Available: https://isei-ivei.euskadi.eus/documents/d/guest/informe-eficaces_cas_def?imagePreview=1
- [10] M. Birenbaum, "Assessment 2000: Towards a Pluralistic Approach to Assessment," in *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge*, 1st ed., M. Birenbaum and F. J. R. C. Dochy, Eds. Dordrecht: Springer Netherlands, 1996, pp. 3–29, doi: 10.1007/978-94-011-0657-3_1.
- [11] L. A. Shepard, "Classroom Assessment to Support Teaching and Learning," *The ANNALS of the American Academy of Political and Social Science*, vol. 683, no. 1, pp. 183–200, May 2019, doi: 10.1177/0002716219843818.
- [12] A. M. Kleinsasser, "Assessment Culture and National Testing," *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, vol. 68, no. 4, pp. 205–210, Apr. 1995, doi: 10.1080/00098655.1995.9957233.
- [13] D. Wolf, J. Bixby, J. Glenn, and H. Gardner, "Chapter 2: To Use Their Minds Well: Investigating New Forms of Student Assessment," *Review of Research in Education*, vol. 17, no. 1, pp. 31–74, Jan. 1991, doi: 10.3102/0091732X017001031.
- [14] P. Black, C. Harrison, C. Lee, B. Marshall, and D. Wiliam, "Working inside the Black Box: Assessment for Learning in the Classroom," *Phi Delta Kappan*, vol. 86, no. 1, pp. 8–21, Sep. 2004, doi: 10.1177/003172170408600105.
- [15] V. L. Mills and C. Harrison, "Intentional Professional Learning Design: Models, Tools, and the Synergies they Produce Supporting Teacher Growth," *Educational Assessment*, vol. 25, no. 4, pp. 331–354, 2020, doi: 10.1080/10627197.2020.1766961.
- [16] F. M. Van der Kleij, J. A. Vermeulen, K. Schildkamp, and T. J. H. M. Eggen, "Integrating data-based decision making, assessment for Learning and diagnostic testing in formative assessment," *Assessment in Education: Principles, Policy & Practice*, vol. 22, no. 3, pp. 324–343, Jul. 2015, doi: 10.1080/0969594X.2014.999024.
- [17] L. A. Shepard, W. R. Penuel, and K. L. Davidson, "Design principles for new systems of assessment," *Phi Delta Kappan*, vol. 98, no. 6, pp. 47–52, Mar. 2017, doi: 10.1177/0031721717696478.
- [18] K. Schildkamp, F. M. van der Kleij, M. C. Heitink, W. B. Kippers, and B. P. Veldkamp, "Formative assessment: a systematic review of critical teacher prerequisites for classroom practice," *International Journal of Educational Research*, vol. 103, p. 101602, 2020, doi: 10.1016/j.ijer.2020.101602.
- [19] M. Heritage, "Getting the Emphasis Right: Formative Assessment through Professional Learning," *Educational Assessment*, vol. 25, no. 4, pp. 355–358, Oct. 2020, doi: 10.1080/10627197.2020.1766959.
- [20] P. Black and D. Wiliam, "Classroom assessment and pedagogy," *Assessment in Education: Principles, Policy & Practice*, vol. 25, no. 6, pp. 551–575, Nov. 2018, doi: 10.1080/0969594X.2018.1441807.
- [21] L. Shepard, "Ambitious Teaching and Equitable Assessment: A Vision for Prioritizing Learning, Not Testing," *American Educator*, vol. 45, no. 3, pp. 28–37, 2021.
- [22] L. Allal, "The Co-Regulation of Student Learning in an Assessment for Learning Culture," in *Assessment for Learning: Meeting the Challenge of Implementation*, D. Laveault and L. Allal, Eds. Cham: Springer International Publishing, 2016, pp. 259–273, doi: 10.1007/978-3-319-39211-0_15.
- [23] L. Adie, B. Addison, and B. Lingard, "Assessment and learning: an in-depth analysis of change in one school's assessment culture," *Oxford Review of Education*, vol. 47, no. 3, pp. 404–422, May 2021, doi: 10.1080/03054985.2020.1850436.
- [24] L. Kyriakides, B. P. M. Creemers, A. Panayiotou, and E. Charalambous, *Quality and Equity in Education*. London: Routledge, 2020, doi: 10.4324/9780203732250.
- [25] J. Willis, K. McGraw, and L. Graham, "Conditions that mediate teacher agency during assessment reform," *English Teaching: Practice & Critique*, vol. 18, no. 2, pp. 233–248, Jun. 2019, doi: 10.1108/ETPC-11-2018-0108.
- [26] C. DeLuca, A. Chapman-Chin, and D. A. Klinger, "Toward a Teacher Professional Learning Continuum in Assessment for Learning," *Educational Assessment*, vol. 24, no. 4, pp. 267–285, Oct. 2019, doi: 10.1080/10627197.2019.1670056.
- [27] M. Grangeat, C. Harrison, and J. Dolin, "Exploring assessment in STEM inquiry learning classrooms," *International Journal of Science Education*, vol. 43, no. 3, pp. 345–361, Feb. 2021, doi: 10.1080/09500693.2021.1903617.
- [28] S. Doğan and A. Adams, "Effect of professional learning communities on teachers and students: reporting updated results and raising questions about research design," *School Effectiveness and School Improvement*, vol. 29, no. 4, pp. 634–659, Oct. 2018, doi: 10.1080/09243453.2018.1500921.
- [29] C. Wolterinck, C. Poortman, K. Schildkamp, and A. Visscher, "Assessment for Learning: Developing the required teacher competencies," *European Journal of Teacher Education*, vol. 47, no. 4, pp. 711–729, Aug. 2024, doi: 10.1080/02619768.2022.2124912.
- [30] J. Feldman, "The role of professional learning communities to support teacher development: a social practice theory perspective," *South African Journal of Education*, vol. 40, no. 1, pp. 1–8, Feb. 2020, doi: 10.15700/saje.v40n1a1668.




- [31] M. Christoforidou and L. Kyriakides, "Developing teacher assessment skills: the impact of the dynamic approach to teacher professional development," *Studies in Educational Evaluation*, vol. 70, p. 101051, Sep. 2021, doi: 10.1016/j.stueduc.2021.101051.
- [32] J. A. de Vries, A. Dimosthenous, K. Schildkamp, and A. J. Visscher, "The impact on student achievement of an assessment for learning teacher professional development program," *Studies in Educational Evaluation*, vol. 74, p. 101184, Sep. 2022, doi: 10.1016/j.stueduc.2022.101184.
- [33] M. Birenbaum, "Conceptualizing Assessment Culture in School," in *Designing Assessment for Quality Learning*, C. Wyatt-Smith, V. Klenowski, and P. Colbert, Eds. Dordrecht: Springer Netherlands, 2014, pp. 285–302, doi: 10.1007/978-94-007-5902-2_18.
- [34] M. Birenbaum, "Assessment Culture Versus Testing Culture: The Impact on Assessment for Learning," in *Assessment for Learning: Meeting the Challenge of Implementation*, D. Laveault and L. Allal, Eds. Cham: Springer International Publishing, 2016, pp. 275–292, doi: 10.1007/978-3-319-39211-0_16.
- [35] M. Birenbaum, H. Kimron, H. Shilton, and R. Shahaf-Barzilay, "Cycles of inquiry: formative assessment in service of learning in classrooms and in school-based professional communities," *Studies in Educational Evaluation*, vol. 35, no. 4, pp. 130–149, Dec. 2009, doi: 10.1016/j.stueduc.2010.01.001.
- [36] M. Birenbaum, H. Kimron, and H. Shilton, "Nested contexts that shape assessment for learning: school-based professional learning community and classroom culture," *Studies in Educational Evaluation*, vol. 37, no. 1, pp. 35–48, Mar. 2011, doi: 10.1016/j.stueduc.2011.04.001.
- [37] Boletín Oficial del Estado Consolidada Legislación, "Organic Law 2/2006, of May 3, on Education," (in Spanish), 2006. [Online]. Available: <https://www.boe.es/buscar/act.php?id=BOE-A-2006-7899>
- [38] Departamento de Educacion "Decree 175/2007, dated October 16, establishing the curriculum for Basic Education and implementing it in the Autonomous Community of the Basque Country. (BOPV No. 218, dated November 13, 2007)," (in Spanish), 2007. [Online]. Available: <https://www.euskadi.eus/web01-bopv/es/p43aBOPVWebWar/VerParalelo.do?cd2007006182>
- [39] Basque Government, *Diagnostic assessment in the Basque Country: a proposal for its development and implementation*. ISEI-IVEI (in Spanish), 2008. [Online]. Available: https://www.euskadi.eus/contenidos/documentacion/inn_doc_comp_basicas/es_def_adjuntos/evaluacion/311002c_Pub_ISEI_marco_evaluacion_diagnostica_c.pdf
- [40] J. W. Creswell, *A Concise Introduction to Mixed Methods Research*, 2nd ed. Thousand Oaks, CA: SAGE Publications, Inc., 2021.
- [41] R. K. Yin, *Case Study Research and Applications: Design and Methods*, 6th ed. Thousand Oaks, CA: SAGE Publications, Inc., 2018.
- [42] P. Mayring, *Qualitative Content Analysis Theoretical Foundation, Basic Procedures and Software Solution*. Klagenfurt: SSOAR, 2014.
- [43] L. Joaristi, L. Lizasoain, and V. Azpillaga, "Detection and characterization of highly effective schools in the autonomous Community of the Basque Country using contextualized cross-sectional attainment models and hierarchical linear models," (in Spanish), *Estudios sobre Educación*, vol. 27, pp. 37–61, 2014, doi: 10.15581/004.27.37-61.
- [44] H. T. Gebremariam and A. D. Gedamu, "Primary school teachers' assessment for learning practice for students' learning improvement," *Frontiers in Education*, vol. 8, p. 1145195, Apr. 2023, doi: 10.3389/educ.2023.1145195.
- [45] A. Kanjee, "Exploring primary school teachers' use of formative assessment across fee and no-fee schools," *South African Journal of Childhood Education*, vol. 10, no. 1, pp. 1–13, Jun. 2020, doi: 10.4102/sajce.v10i1.824.
- [46] H. Fives and N. Barnes, "Navigating the complex cognitive task of classroom assessment," *Teaching and Teacher Education*, vol. 92, p. 103063, Jun. 2020, doi: 10.1016/j.tate.2020.103063.
- [47] M. Mohamed, M. S. A. Aziz, and K. Ismail, "'Assessment for Learning' Practices Amongst the Primary School English Language Teachers: A Mixed Methods Approach," *Pertanika Journal of Social Sciences and Humanities*, vol. 29, no. 3, pp. 1875–1900, Sep. 2021, doi: 10.47836/pjssh.29.3.21.
- [48] P. Mathew, P. Mathew, and J. Peechattu, "Reflective Practices: a Means To Teacher Development," *Asia Pacific Journal of Contemporary Education and Communication Technology (APJCECT)*, vol. 3, no. 1, pp. 126–131, 2017.
- [49] A. Hargreaves, "Teacher collaboration: 30 years of research on its nature, forms, limitations and effects," *Teachers and Teaching*, vol. 25, no. 5, pp. 603–621, Jul. 2019, doi: 10.1080/13540602.2019.1639499.
- [50] L. Stoll, L. Earl, A. Stephen, and K. Schildkamp, "Educational effectiveness and improvement research, and teachers and teaching," in *The Routledge International Handbook of Educational Effectiveness and Improvement: Research, policy, and practice*, C. Chapman, D. Muijs, D. Reynolds, P. Sammons, and C. Teddlie, Eds. London: Routledge, 2015, pp. 380–396, doi: 10.4324/9781315679488-23.
- [51] K. Kager, J. P. Mynott, and M. Vock, "A conceptual model for teachers' continuous professional development through lesson study: capturing inputs, processes, and outcomes," *International Journal of Educational Research Open*, vol. 5, p. 100272, Dec. 2023, doi: 10.1016/j.ijedro.2023.100272.
- [52] H.-L. W. Pan and S.-H. Cheng, "Examining the impact of teacher learning communities on self-efficacy and professional learning: an application of the theory-driven evaluation," *Sustainability*, vol. 15, no. 6, p. 4771, Mar. 2023, doi: 10.3390/su15064771.

BIOGRAPHIES OF AUTHORS






Izaskun Lopetegi-Mendizabal    is a lecturer and Ph.D. student at the Department of Education Sciences, within the Faculty of Education, Philosophy and Anthropology at the University of the Basque Country (UPV/EHU). Her research mainly focuses on assessment culture in education. She has an undergraduate degree in pedagogy and a university master's degree in teaching training, specializing in educational guidance, from the University of the Basque Country (UPV/EHU). She is a member of the GANDERE (gender socialization in educational contexts) group, which has been officially recognized by the UPV/EHU. She is currently working towards earning a Ph.D. in Education: school, language and society. She can be contacted at email: izaskun.lopetegi@ehu.eus.



Verónica Azpillaga-Larrea    is an associate lecturer at the Department of Education Sciences, within the Faculty of Education, Philosophy and Anthropology at the University of the Basque Country (UPV/EHU). She has worked as a guidance counselor at several schools in the public and semi-private education systems, at both primary and secondary level. In the university field, she has lectured on both undergraduate (pedagogy, social education) and postgraduate courses (Master's in Secondary Education, European Master's in Education and Multilingualism), and her work focuses mainly on issues linked to psycho-pedagogical guidance and the pedagogy of language and communication. Since 2013, she has participated as a researcher in projects associated with School Effectiveness and Improvement in the Autonomous Community of the Basque Country. She is a member of the GANDERE (gender socialization in educational contexts) group, which has been officially recognized by the UPV/EHU. She can be contacted at email: veronica.azpillaga@ehu.eus.



Luis Lizasoain Hernández    has been a lecturer in Education Research Methods at the Faculty of Education, Philosophy and Anthropology at the University of the Basque Country (UPV/EHU) since January 2019. His research mainly focuses on the evaluation of programmers, schools and education systems, with special emphasis on large-scale educational assessments and school effectiveness and improvement studies, alongside the statistical data analysis techniques associated with them. He has a special interest in statistics decision trees and multilevel regression procedures using hierarchical linear models. He has coordinated a consolidated research team at the UPV/EHU and has been lead researcher in several projects on school effectiveness and improvement in the Basque Country. His scientific work has led to numerous publications in high-impact journals. He can be contacted by email at: luis.lizasoain@ehu.eus.