European Journal of Psychology of Education (2025) 40:55
https://doi.org/10.1007/s10212-025-00951-9

How dominant language influences rubric reading and task performance: Insights from eye-tracking research

Ernesto Panadero[1,2,3] · Pablo Delgado[4] · Lucía Barrenetxea-Mínguez[2] · David Zamorano[2] · Leire Pinedo[2] · Alazne Fernández-Ortube[2]

Received: 15 July 2024 / Revised: 11 December 2024 / Accepted: 28 February 2025

Abstract

The students' dominant language might influence how they use and process a rubric and its subsequent effect on task performance. However, our knowledge about these effects is limited. This study investigates how the dominant language of students is associated with their rubric reading patterns and their task performance in a written landscape analysis in Spanish. Participants were 80 higher education students with different dominant language (Spanish-dominant speakers, SDS; Basque-Spanish speakers, BSS) from six undergraduate programmes. We employed a randomized controlled trial in which participants used a rubric to guide their performance in a written analysis of a landscape. Participants were randomly assigned to two conditions based on the rubric order: (1) lowest to highest performance level vs (2) highest to lowest performance level. We analyzed eye-tracking data to explore reading patterns (i.e., fixation times on the rubric cells and gaze transitions between the rubric and the picture of the landscape), task performance (i.e., written landscape analysis), and self-reported cognitive load. Spanish-dominant speakers exhibited more adaptive reading patterns and performed better in the written landscape analysis with the highest-lowest performance level (PL) order rubric, compared to Basque-Spanish speakers. Additionally, fixation time on highest PL and gaze transitions between highest PL and the landscape picture were positively correlated with task performance. Our research highlights the importance of considering dominant language in rubric design and implementation, showing that strategic rubric design can enhance student performance, particularly in linguistically diverse educational settings.

Highlights

- Dominant language affects students' engagement with rubrics and task performance.
- Spanish-dominant speakers performed better with a highest to lowest (PL4–PL1) performance level order rubric.
- Eye-tracking revealed that fixation on PL4 and transitions to the landscape image improved performance.
- Strategic rubric design enhances student engagement and performance in diverse settings.
- Adapting rubrics to linguistic backgrounds can create more equitable educational tools.

Extended author information available on the last page of the article

Published online: 27 March 2025

Keywords Rubric · Feedback · Eye-tracking · Bilingualism · Reading · Writing

Introduction

In academic settings, the dominant language of students emerges as a crucial factor in shaping their educational outcomes, especially when the language of instruction differs from their dominant language. Research has consistently demonstrated that mastery of the academic vocabulary of the instruction language can significantly predict and enhance school performance (Schuth et al., 2017). However, the influence of language goes beyond mere vocabulary proficiency; it also includes the students' frequency of language use and their overall comfort with academic discourse (Treffers-Daller, 2019). This complex interplay between language proficiency, language use, and familiarity with academic language points to a nuanced understanding of how language dominance impacts academic achievement. As such, educators and researchers must consider these linguistic dimensions to develop more effective teaching strategies, educational tools, and improve educational assessments (Leibowitz, 2005).

When it comes to educational assessment, rubrics are widely used around the world (Dawson, 2017; Panadero et al., 2023). Rubrics can enhance students' performance when used as learning and instructional instruments (Panadero et al., 2023). One of the ways in which rubrics are being used is in relationship to self-assessment (e.g., Brookhart, 2018) or to generate self-feedback (Lipnevich et al., 2022; Peltzer et al., 2024). However, how students read and work with rubrics has received very little empirical attention. Given the prevalent use of rubrics and the limited insights into how students use them, it is essential to reevaluate if they are designed and applied effectively to match the ways students work with them.

As rubrics are essential tools for providing feedback and enhancing students' performance, this study investigates how the dominant language of students is associated with their rubric reading patterns and their task performance in a written landscape analysis. In this study, participants, who were first-year university students, performed a landscape analysis task as part of an educational assessment. The rubric, presented in Spanish, was designed to guide students' written analysis by clearly outlining the quality expectations across performance levels. This study examines how participants, divided into Spanish-dominant and Basque-Spanish bilingual groups, used the rubric under different design conditions (highest to lowest -PL4-PL1- vs. lowest to highest -PL1-PL4- performance level order) and how their linguistic background was associated with their reading patterns and task performance. To achieve this, we employed a mixed-methods approach, combining eye-tracking measures with task performance (i.e.,written landscape analysis) to explore these interactions. By utilizing eye-tracking technology, this research offers empirical insights into how students read and interact with rubrics, particularly in multilingual settings.

Dominant language in educational settings and its effects

In educational settings, understanding the concept of *dominant language* is essential, particularly in environments rich in linguistic diversity such as regions with multiple official languages (e.g., Cenoz, 2009). Defined primarily by a person's relative proficiency across

their languages, dominant language indicates which language an individual can use more fluently and effectively. For example, Treffers-Daller (2019) notes that language dominance involves both a person's proficiency in specific language skills, such as vocabulary and grammar, and the extent to which they use these languages in various life domains. This dual focus highlights how dominant language not only concerns the ability to use language but also the frequency and context of its use (Treffers-Daller, 2019). To account for this, we have explored three areas that allows us to consider not only the ability, but also the frequency and context of its use: (1) self-perceived dominant language, (2) mother tongue, and (3) dominant language of instruction during previous education.

Extending this perspective, Aronin (2016) introduces the concept of *dominant language constellations*, which redefines linguistic dominance by emphasizing functional multilingualism where dominant languages collectively meet communicative needs in diverse settings (Aronin, 2016). Additionally, Lasekan (2020) explores how interpersonal relationships influence language dominance in oral and written communication among multilinguals in southern India, suggesting that social interactions play a crucial role in shaping language use (Lasekan, 2020). Moreover, Poeste et al. (2019) investigate the impact of code-mixing on language dominance in multilingual children, providing insights into how linguistic dominance manifests in young learners' language behaviors.

The impact of dominant language proficiency on academic outcomes is well-documented across various educational settings. For example, studies have shown that students with higher language proficiency in the dominant language of instruction exhibit greater academic self-efficacy, which in turn positively affects their academic performance and reduces tendencies like procrastination (Lowinger et al., 2014). Zhuang (2023) highlights that various English language learning methods positively impact students' self-motivation, mediated by their self-efficacy, suggesting that mastering the dominant language not only improves direct academic performance but also enhances students' engagement and intrinsic motivation to learn (Zhuang, 2023). In our study, we will explore how the dominant language moderates how the order of presentation of performance level criteria (i.e., from lowest to highest and vice versa) affects students' reading and use of a rubric, and what is the effect on task performance.

Dominant language as an individual variable moderating rubrics' effect

Like many other elements related to educational assessment, individual differences can play a role in how students process and respond to feedback situations (Panadero, 2023). It can also be expected that the use of rubrics would be associated with students' characteristics, an aspect which remains understudied. Nevertheless, the use of rubrics for formative purposes is often intended to enhance students' self-assessment (Andrade et al., 2010; Lipnevich et al., 2022; Peltzer et al., 2024). Importantly, research has shown how individual differences can be associated with self-assessment (Yan, 2018; Yan et al., 2020). Therefore, it can be expected that students' characteristics may also moderate their use of rubrics. One that could particularly affect is whether students work in their dominant language.

To our knowledge, no previous research has addressed this specific topic, but there is existing research on differences between monolingual and bilingual students focused on executive function skills. Some of this executive function skills might be related to performance in rubric use during self-regulated tasks, such as those related to shifting between different informational stimuli (e.g., Bialystok & Martin, 2004), the inhibition of irrelevant

information (e.g., Martin-Rhee & Bialystok, 2008), or working memory capacity (e.g., Carlson & Meltzoff, 2008). However, as the variable we investigate here is different in nature—i.e., dominant language—, we will operate on the basis that there is not previous empirical evidence on whether students perform better in their dominant language when contrasted with others with a different dominant language. To fill this gap, the present study contributes by also exploring possible differences between Spanish-dominant speakers vs. Basque-Spanish speakers when reading and using a rubric written in Spanish.

Rubrics: definition, educational effects, and theoretical underpinnings

A rubric can be defined as a tool that "articulates expectations for student work by listing criteria for the work and performance level descriptions across a continuum of quality" (Brookhart, 2018, p. 1). Rubrics are commonly designed as tables or matrices, with the first column listing the assessment criteria and subsequent columns describing levels of performance, which can range from high to low quality or vice versa (Panadero et al., 2023).

Rubrics were originally created for summative purposes, but they can currently serve both summative and formative functions (e.g., Brookhart, 2018). When the goal is to have students work with them, they are often used with formative purposes (Panadero & Alonso-Tapia, 2013). Research has shown that such implementation can improve academic performance and planning of assignments, and reduce anxiety and negative self-regulation (Brookhart & Chen, 2015; Panadero et al., 2023). In this line of work, the use of rubrics has been explored extensively for their effectiveness in providing structured and criteria-based feedback to students. For instance, Steiss et al. (2024) compared the quality of feedback provided by human evaluators and ChatGPT—an AI tool—using rubrics. Their findings underscore the importance of rubrics in ensuring high-quality feedback, as human feedback was generally more effective in all categories except for being criteria-based. While rubrics are widely praised for their clarity and structured feedback, some scholars have raised concerns (for a review see Panadero & Jonsson, 2020). Among others, critics argue that rubrics may constrain creativity, fostering a mechanistic adherence to criteria at the expense of deeper, more critical engagement with the task (e.g., Torrance, 2007). Nevertheless, when it comes to the strength of empirical evidence on the positive vs. negative effects of rubrics, there is much stronger evidence on the positive outcomes (Panadero & Jonsson, 2020).

Rubrics enhance self-regulated learning by supporting goal-setting, monitoring progress, and fostering self-reflection. Clearly articulating expectations and providing structured performance criteria, rubrics empower students to independently assess and adjust their work (Panadero & Jonsson, 2013). This study designed a rubric to guide students' in task performance, enabling them to plan, evaluate, and refine their written analysis according to the highest performance standards. Further reinforcing this concept, Panadero and Jonsson (2013) proposed a model suggesting that rubrics clarify expectations, which in turn fosters self-regulated learning and positively influences academic performance. By offering detailed criteria, rubrics help students self-assess and identify areas for improvement, thereby contributing significantly to learning by supporting them in becoming more autonomous and reflective learners. This capability to generate their own feedback based on the rubric is supported by previous research (Andrade et al., 2010; Lipnevich et al., 2022; Peltzer et al., 2024).

Importantly, the theoretical underpinnings of feedback in educational settings, as articulated by Hattie and Timperley (2007), Narciss (2008), and Panadero and Lipnevich (2022), provide a robust framework for understanding how rubrics function as self-feedback tools. Hattie and Timperley's (2007) model, which emphasizes the importance of feedback addressing three key questions (Where am I going? How am I going? Where to next?) at four different levels (task, process, self-regulation, and information about the self), can be directly applied to rubrics. Rubrics offer students clear criteria and performance levels, helping them understand their current performance and identify specific areas for improvement, thereby aligning with the "feed up" (in correspondence with "Where am I going?"), "feedback" ("How am I going?"), and "feedforward" ("Where to next?") components of Hattie and Timperley's model. Similarly, Narciss' Interactive Tutoring Feedback Model (ITF) highlights the need for feedback to be specific, timely, and detailed. Rubrics, by providing detailed performance criteria and quality levels, enable learners to adjust their cognitive processes, correct misconceptions, and build upon their existing knowledge in a structured manner. Furthermore, Panadero and Lipnevich's (2022) synthesis of feedback models underscores the multifaceted nature of feedback and its role in fostering self-regulated learning. Their MISCA model's focus on feedback message, implementation, and student characteristics is mirrored in the use of rubrics, which help students generate their own feedback, implement changes, and reflect on their performance. Thus, these theoretical frameworks collectively support the use of rubrics as effective self-feedback tools that promote autonomous and reflective learning.

Student use of rubrics: reading and writing

Understanding how students read and use rubrics is essential for enhancing their formative impact (e.g., Andrade & Du,2007; Turley & Gallagher, 2008). While research has explored how raters and teachers employ rubrics (e.g., Postmes et al., 2023; Winke et al., 2015), much less attention has been given to how students engage with them. Some studies have examined students' retrospective perceptions of rubric use (e.g., Andrade& Du, 2005), but only a few, to our knowledge, have directly analyzed students' interactions with rubrics using process data (e.g., Panadero et al.,2012). Process data, which includes methods such as observations, think-aloud protocols, and eye-tracking (Panadero, 2023), provides a more objective and real-time understanding of students' engagement with rubrics. Among these methods, eye tracking is particularly valuable for examining how students navigate and interpret rubric information in real time.

Importantly, there is limited research using eye-tracking to study rubric use beyond students. Winke & Lim (2015) observed that professional raters primarily focused on the first columns of a rubric when assessing essays, although these findings might be biased due to the non-counterbalanced order of criteria across the columns. This underscores the need for more robust empirical studies involving students. In the present study, we will use eye-tracking data to investigate how students read a rubric written in Spanish in a context with linguistically diverse students (Spanish-dominant vs Basque-Spanish speakers).

Understanding how students read rubrics is crucial for enhancing their effectiveness as formative assessment tools. When students read rubrics, they are engaging in a process that involves decoding the criteria and performance levels to understand the expectations for their work (Andrade & Du, 2007). This process is informed by theoretical frameworks that emphasize the importance of clarity and specificity in feedback (Hattie & Timperley, 2007). Research suggests that students often focus on specific performance

levels or criteria that they find most relevant or challenging (Reddy & Andrade, 2010). Eye-tracking studies, such as those by Winke & Lim (2015), indicate that understanding which parts of the rubric students focus on can reveal important insights into how they interpret and prioritize the feedback provided. The theoretical foundation here is that effective reading of rubrics facilitates self-regulated learning by enabling students to internalize the assessment criteria and use them to guide their performance and to foster improvement (Panadero & Alonso-Tapia, 2013). We understand for "effective reading" the way students processed and prioritized information from the rubric, focusing on the performance level descriptors that guided their writing to improve their work.

Enhancing student writing through rubrics has been widely supported by research (e.g., Peltzer et al., 2024). Rubrics provide a structured framework that helps students understand what constitutes quality work, thereby guiding their writing process (Brookhart & Chen, 2015). The theoretical basis for this is, again, that rubrics clarify expectations and provide detailed descriptions of what is required to achieve different levels of performance, which can help students organize their thoughts, structure their writing, and focus on key aspects of their assignments (Peltzer et al., 2024). This structured approach helps students produce high-quality work by allowing them to self-assess and revise their writing according to the criteria laid out in the rubric.

Rubric layout: design considerations for performance levels

The design of rubrics is one of the key elements that could influence their educational effects (Brookhart, 2018). As mentioned earlier, rubrics are usually structured as tables that vary in complexity based on content organization and language clarity (Brookhart, 2018). If the language is too technical, students may struggle to understand the rubric. Additionally, the number of performance levels needs careful consideration; too many can complicate the rubric's use, while too few may not adequately represent the range of performances. However, research on how rubric design influences its use is scant, with limited studies focusing on teachers rather than students (Humphry & Heldsinger, 2014).

A critical design feature that could significantly impact student interaction with rubrics is the order of performance levels in the rubric. Specifically, the placement of the highest performance level, which typically contains the most detail, may affect how students interact with the rubric. This study will also explore whether students with different dominant language status focus primarily on the highest level and how their interactions vary with the positioning of this level by experimentally manipulating the order from highest to lowest and vice versa. Crucially, the rubric in this study was employed as a formative tool for guiding students in their landscape analysis tasks. Students used the rubric to understand the assessment criteria and align their work with these standards throughout both the analysis of a landscape and the subsequent writing process. Unlike scenarios involving pre-existing feedback or assessments, students were expected to actively interpret and apply the rubric to their own work, using it to guide their decision-making and enhance the quality of their written analysis.

The present study: aim, research questions, and hypotheses

Thus, we aim to investigate whether students' dominant language moderates the potential effect of the order of performance level criteria on how students read rubrics and their task

performance in a written landscape analysis. For this, we used eye-tracking methods and manipulated the performance levels order of presentation (highest to lowest vs. lowest to highest). We explored two research questions (RQ):

RQ1. Does students' dominant language moderate differences in rubric reading patterns based on the order of presentation of the rubric performance levels?
RQ2. How is students' dominant language associated with their task performance (written landscape analysis) when using rubrics?

Given the non-existence of research on the effects of dominant language on the use of rubrics, we proposed the following general non-directional hypotheses: (H1) Basque-Spanish speakers will show different reading patterns of the rubric than Spanish dominant speakers across different rubric designs, which will impact their task performance in a written landscape analysis (H2). Additionally, (H3) presenting the highest performance level first (PL4) will direct more attention towards the highest standard, influencing reading patterns and possibly performance.

Importantly, the operationalization of language dominance can be assessed by employing both direct measures, such as proficiency tests, and indirect methods, like exposure and usage surveys. We chose to survey participants, university students who can reliably report their language use, believing this method captures the necessary data without the complexity and burden of proficiency tests, which could potentially decrease participant engagement or fidelity in our study.

Method

Experimental design

We implemented an experimental design that included a between-participant experimental factor—the order of performance levels (PL) in the rubric—and a potential moderator—students' dominant language. Rubric PL featured four levels (PL1 to PL4, from lowest to highest, or viceversa). Participants were randomly assigned to one of two orders of rubric PL: either ascending from PL1 to PL4 or descending from PL4 to PL1 (always from left to right as that is the reading direction in Spanish). Additionally, we examined participants' dominant language by dividing the participants in two groups: (1) Spanish-dominant speakers and (2) Basque-Spanish speakers (see Participants section for a detailed description). The order of the rubric performance levels was manipulated (PL1–PL4 vs. PL4–PL1) to explore the potential effect of the performance level order on students' reading patterns and task performance. This manipulation was grounded in cognitive load theory (Sweller, 2010) and primacy effects, which suggest that the sequence in which information is presented can significantly shape how it is processed and used. By placing the most detailed and informative level (PL4) first, we hypothesized that participants might focus more on the highest performance standard. Our experimental design also builds on findings from Winke & Lim (2015), who observed that raters often focus on the first column of a rubric, so we aimed to test whether this pattern extends to students in the context of formative assessment.

Participants

Inclusion and exclusion criteria

We recruited students from six distinct undergraduate programs at the same university. The sole exclusion criterion was having vision impairments or undergoing recent surgical procedures. Only one participant was excluded based on these criteria. All other participants possessed normal or corrected-to-normal vision.

Territorial and educational context of the study

The study was conducted in the Basque Autonomous Region (Spain) with two official languages: Basque and Spanish. As reported by Cenoz (2009), bilingual and multilingual education in the Basque Autonomous Region is of international interest due to several unique aspects. Firstly, the revival of Basque as a minority language for instruction in combination with Spanish and other foreign languages presents a complex scenario of language learning and heritage language preservation. This is evident not only among native Basque speakers but also among those who learn it as a second language at school. Secondly, the categorization of Basque as a heritage, second, or foreign language is challenging because it does not neatly fit into these definitions. While Basque has received very substantial institutional support and is widely used in educational and cultural domains, it is not the dominant language in terms of mother tongue (17.8%) or language spoken at home (13.1%) across the entire population of the Basque Autonomous Region (Eustat, 2021). This creates a unique linguistic context, where students may not use Basque as their primary language in everyday tasks despite its official status in educational context where is clearly the main language due to educational policies.

Additionally, the Basque educational system blurs traditional distinctions in bilingual education such as elite versus folk bilingualism. Despite significant investment, it is not elitist as it targets all students, and speakers of Basque are not economically disadvantaged. The system extends from kindergarten through university, highlighting the evolving role of Basque and the varied challenges at different educational levels. This complexity illustrates the difficulty of strictly categorizing linguistic and educational dynamics, emphasizing the overlapping areas of heritage, second language, and foreign language education within the region.

Sampling procedure

Our study utilized a convenience sample comprising first-year university students enrolled in a variety of programs—Psychology, Medicine, Education, Sport Sciences, Social Work, and Social Education—from the Universidad de Deusto (Bilbao campus). We recruited participants by contacting their instructors via email to invite the students to a session where the study was introduced. During these sessions, the purpose and importance of the study were explained, and students interested in participating were asked to provide their email addresses on a sign-up sheet. We reached out to seven classroom groups, resulting in 174 students consenting to contact, and 138 of them attended the laboratory sessions. Nevertheless, due to the need to filter out low-quality eye-tracking data (refer to the Apparatus and Measures section for further details), the final sample size was reduced to 80 participants (58% of those who came to the laboratory), including 55 females. Participants

who completed the experiment were compensated with a financial incentive of 5€ for their participation.

Participants' academic background

The final sample ($N=80$) comprised students majoring in Psychology (43.75%), Medicine (18.75%), Education (12.50%), Sport Sciences (12.50%), Social Work (8.75%), and Social Education (3.75%). While our sample was a convenience sample, its diversity across programs contributes to its representativeness of first-year students in the university. However, we acknowledge that this method does not capture all potential students, and our findings should be interpreted within the context of this sampling approach.

The participants had an average pre-university academic achievement of 8.3 out of 10, indicating a high academic standard within the sample. We next report the distribution of the sample regarding each of the three variables related to their language status.

Student self-reported dominant language

After our experimental data collection students self-reported their dominant language (Spanish or Basque) using a 10-point Likert-scale. The value 1 indicated that they only spoke Spanish and value 10 that they only spoke Basque. Value 5 indicated similar proficiency in both languages. We identified as Spanish-dominant speakers (SDS) as those who rated themselves from 1 to 3, and as Basque-Spanish speakers (BSS) if their ratings were above 3. Although students who rated themselves above 7 would strictly be considered as Basque-dominant, only six students met this criterion. Given the small size of this group and the fact that the experiment was conducted entirely in Spanish, we opted to include them within the BSS category. Out of the 80 participants in the final sample, 43 students were considered SDS (53.75%), while 37 (46.25%) were considered BSS (see Table 1). Importantly, we asked participants to report two additional language aspects (i.e., mother tongue and language in compulsory education). However, the distribution of the participants was skewed (see Table 1 with the data of these two examples to visualize the skewness of the distribution), and it was not possible to further explore them.

Table 1 Distribution of values for self-reported dominant language, mother tongue, and language in compulsory education

Language status variable	Values	*n* in 1–4 PL order condition	*n* in 4–1 PL order condition
Self-reported dominant language	Spanish	20	23
	Basque-Spanish	19	18
Mother tongue (not analyzed)	Spanish	23	33
	Basque	12	7
	Spanish-Basque	4	1
Language in compulsory education (not analyzed)	Spanish	7	6
	Basque	28	27
	Trilingual	4	8

To provide clarity on the linguistic context of our participants, it is important to differentiate between primary language, dominant language, and language of instruction. Primary language refers to the language an individual is first exposed to, typically spoken in the home environment, often referred to as the "mother tongue." However, in multilingual regions like the Basque Autonomous Region, the primary language may not always correspond to the language in which individuals are most proficient or comfortable in academic or everyday settings. This leads to the concept of dominant language, which reflects the language an individual uses most frequently and fluently across various life domains. In this study, we focused on self-reported dominant language as it captures the nuanced linguistic reality of bilingual participants more comprehensively than primary language. Finally, language of instruction refers to the language used in educational contexts. For our participants, Spanish was the primary language of instruction, even for those whose dominant language was Basque. In our classification, Basque-Spanish speakers were bilingual individuals who self-reported Basque as either equally or more dominant than Spanish, whereas Spanish-dominant speakers reported higher proficiency and comfort with Spanish. This distinction allows for a deeper understanding of how linguistic backgrounds is associated with rubric use and task performance in a multilingual educational context.

Materials and intervention

Rubric

The rubric, developed in a prior study (Panadero et al., 2012), was informed by frameworks and guidelines established by social science education experts, particularly those emphasizing the role of rubrics in fostering self-regulated learning and structured feedback (Andrade et al., 2010; Panadero & Alonso-Tapia, 2013). It includes four performance levels and five assessment criteria (refer to Appendix B for details). In this study, the order of the performance levels and assessment criteria was manipulated to gather insights into the reading patterns. The rubric was presented in Spanish.

Task: written landscape analysis

The task consisted of performing a written landscape analysis, where participants were required to describe and analyze two distinct rural landscapes: one representing a Continental climate and the other a Mediterranean climate (see Appendix C for details). This type of exercise is rooted in the Spanish secondary education curriculum and is designed to foster observational and analytical skills by focusing on physical features, climate, and land use. Given this curricular background, it can be assumed that participants from diverse university programs would possess the foundational knowledge required for the task. The rubric provided criteria to guide and self-monitor their analyses, ensuring consistency and alignment with the study's objectives.

The landscape was presented (1) alongside the rubric during the oral analysis and (2) again with the rubric and accompanied by a text processor (i.e., a Microsoft Word document) where participants wrote their final analysis of the landscape. This written analysis was used to measure the participants' task performance. A crucial aspect is that all participants performed the landscape analysis in Spanish.

Instructions

All the participants received the same verbal instructions for the task, detailed in Appendix D.

Apparatus and measures

Eye tracker

Stimuli were displayed on a 24-inch screen with a resolution of 1680 × 1050 pixels. Eye movements were tracked using a Tobii Pro Fusion screen-based eye-tracker operating at 250 Hz under Tobii Pro Lab software. The eye tracker was calibrated for optimal accuracy before each participant began performing the task, as a first way to prevent extreme data loss or miscalibration, the calibration was repeated if the participants exceeded the 15% data loss threshold or/and 0.5° of accuracy (Holmqvist et al., 2011). If the results were consistent in the subsequent calibration, the experimental procedure continued and the output was manually corrected in cases of miscalibration, more details below.

A timed calibration approach was employed, whereby calibration points were only displayed once the eye tracker had gathered sufficient data to continue. Additionally, the order of the calibration target points was randomized to prevent anticipatory gaze behaviors. The calibration used a grey background with white targets, and the system's default setting of five calibration points was maintained (Tobiipro, 2021).

This study is part of a larger study where concurrent thinking aloud protocols (CTAP) were performed while recording with the eye tracker, which might have caused data loss (Holmqvist et al., 2011). Therefore, to ensure the quality of the eye tracker measures, participants' data loss was examined through Tobii Pro Lab eye-tracking software, and a threshold was established: 25% of each participant's gaze not sampled by the eye-tracking system, as indicated by the Tobii Pro Lab software. All participants surpassing that threshold were excluded. The raw eye-tracking data of the remining participants ($N = 80$) were visually scanned by the fourth, fifth, and sixth authors to further assess possible miscalibrations along the entire recording caused by the CTAP. They performed four rounds of inter-rater agreements, with 15% of the recordings each, until an 85.3% percentage of agreement was reached.

We measured participants' total fixation time on each rubric cell (i.e., areas of interest) in two testing times: (1) initial reading of the rubric and (2) oral analysis. Those fixations on the rubric's textual content that occurred during participants think-aloud utterances were excluded when computing total fixation times, as it is not possible to process textual information while different speech (i.e., think-aloud utterances) is being produced. The exact times at which participants uttered verbalizations were identified, and fixations that occurred in those time intervals were removed. As the textual information in the rubric varied in length across cells (see Fig. 1), total fixation time on each cell was averaged by the number of characters. Fixations longer than 1000 ms were also removed, as extremely long fixations are deemed to be the product of readers' momentary loses of concentration. Conklin and colleagues (Conklin et al., 2018) recommend excluding fixations longer than 800 ms, whereas other authors opt for a much more conservative threshold of 1200 ms (e.g., Rivero-Contreras et al., 2023). We thus opted for a balanced criterion. Fixations were identified by means of the Tobii I-VT (fixation) algorithm. This pre-set, optimized for screen-based eye-trackers, is the default pre-set

Step 1 Initial reading rubric

Step 2 Oral Landscape analysis: Rubric + Landscape

Step 3 Written Landscape analysis: Rubric + Landscape + Text processor

Fig. 1 A representation of the stimulus

for screen projects (I-VT fixation classifier: default: threshold 30 degrees/second; merge adjacent fixations: default: enabled; max time between fixations: 75 ms; max angle between fixations: 0.5 degrees; discard short fixations: default: enabled. Minimum fixation duration: 60 ms).

In addition, we measured participants' number of gaze transitions between the rubric and the picture of the landscape (see Fig. 1) during the oral landscape analysis (Phase 2).

Task performance: written landscape analysis

Task performance was assessed by averaging the grades from the written analysis that the participants performed twice during the experimental procedure. Note that we are only using data from the first written landscape analysis in this study. These grades were determined using the established rubric described in Appendix B. To ensure consistency and reliability in scoring, the third and sixth author independently evaluated 15% of the landscape analyses in two rounds, achieving excellent inter-rater reliability (ICC = 0.99; Hallgren, 2012). Any scoring discrepancies were collaboratively resolved by the two coders. The internal consistency of the five rubric items (i.e., assessment criteria) was measured using the omega coefficient (McDonald, 1999), which is preferred over Cronbach's alpha for evaluating items with fewer than five response options (Trizano-Hermosilla & Alvarado, 2016). The omega value was 0.603. Considering that each rubric item referred to different dimensions of the landscape analysis, the reliability can be considered as satisfactory. In Appendix E we present written samples of varying quality from three participants and how we scored them using the rubric.

Students' university entry scores

Students provided their scores from the university entrance (called EBAU). These are calculated by weighting 40% for the university entry exam score and 60% for the final baccalaureate program grade compounded of the last two years of high school training (called Bachillerato).

Additional measures not analyzed in the present study

The data presented in this study belongs to a broader research project. Beyond the instruments and measures previously discussed, all participants were recorded via video and audio during the experimental data collection. This was done to investigate their verbal and emotional responses (i.e., think-aloud protocols) and their actions. Additionally, participants completed various questionnaires to report and control other individual differences. These data fall outside the scope of the current study and are thus not analyzed here. For reasons of brevity, the descriptions of these questionnaires are available in Appendix F.

Procedure

Participants individually attended the laboratory setting accompanied by a member of the research team. All experimental materials and procedures were carried out in Spanish. Before starting the experiment, each student signed an informed consent form detailing the study's purpose and methods. Participants were then briefed on the experimental procedure, including training on how to perform think-aloud tasks, followed by a practice session. After this briefing, participants completed the set of questionnaires.

The experimental setup included a room divider, behind which participants were seated in front of a screen equipped with the eye-tracker, which was calibrated for each participant. The experimental procedure that we will explore in this paper was divided into three phases, though the original procedure expanded two further phases that will not be presented here.

In Phase 1, participants were presented with a rubric to read and familiarize themselves without any further instructions on how to use, allowing for a natural interaction of the participants with the rubric. All participants had prior experience with rubrics. During this time, they were instructed to verbalize their thoughts, feelings, and actions. These think-aloud protocols were recorded as part of a broader research project aimed at understanding students' cognitive and emotional processes while using rubrics. While the present study focuses on eye-tracking data and task performance outcomes, the analysis of verbalizations is addressed in a separate paper.

In Phase 2, the landscape (a rural area with a continental climate) was displayed above the rubric. We asked the participants to analyze the landscape out loud while continuing thinking aloud their thoughts, emotions, etc. Phase 2 finalized once the participant had nothing further to add. Subsequently, Phase 3 started, in which, a word file was displayed alongside the rubric and the landscape. In this phase, we asked participants to write in the word file their final analysis of the landscape they had just orally described, while thinking aloud. Since this paper focuses solely on the data collected up to this point, we do not present the subsequent phases, as they are not relevant to our results. Nevertheless, Appendix G present a figure with the whole experimental procedure for consultation.

Importantly, our study makes a formative use of rubrics. Participants did not receive any scores during the study; instead, they used the rubric to guide and self-assess their task performance (for more details on the implementation of the rubric see Appendix A).

Data analyses

We performed a linear mixed-effect model (LMM) analysis to examine students' fixation times on the rubric across dominant language groups and rubric PL order conditions. The model included fixation time as dependent variable, including the interaction term of rubric PL order and cell PL. Participant language status and rubric cell were included as random effects. This model was constructed as follows. We first built the null model, namely, the model including only the random effect terms. Intercepts were set as random, whereas decisions on the nature of the slopes (i.e., fixed or random) were made based on null models' goodness of fit comparison. The results showed that none of the null models had better goodness of fit than the one including fixed slopes for both random effects. Then, we constructed the final model by adding the interaction term of the fixed factors. Note that although the construction of the null model was statistically driven based on model fit comparisons, the inclusion of the interaction term of the fixed effects was guided by our research questions.

In addition, we performed two generalized linear models (GLM) to examine students' gaze transitions between rubric and the picture of the landscape and students' performance in the landscape analysis task, respectively. In both cases, we included the interaction term of student dominant language and rubric PL order.

Students' university entry scores did not show to be associated with any of the dependent variables. This measure was therefore not included as covariate in any of the models above.

All the analyses were performed using the R software version 4.3.0 in RStudio 2023.03.0 + 386. The LMM was run using the "lmer" function ("lmer4" package v1.1–27; Bates et al., 2021), null-models' goodness of fit was tested with the "anova" function ("stats" package v4.4.0; R Core Team, 2023), confidence intervals of the estimates were calculated using the "confint" function ("stats" package v4.0.2), and models' fit were examined using the "r.squaredGLMM" function ("MuMIn" package v1.47.5; Barto, 2023). Pairwise comparisons of the models' estimate between categorical values of the fixed effects were performed with the "emmeans" function ("emmeans" package v.2.30–0; Lenth et al., 2023). Given that data distribution from fixation times and rubric-picture gaze transitions were not normal, they were log_{10}-transformed using the "log10" function ("base" package v4.3.0; R Core Team, 2023). GLMs were run using the "glm" function ("stats" package v4.4.0). All the syntaxes and the dataset can be *found* at OSF < link TBA > .

Results

RQ1: Does students' dominant language moderate differences in rubric reading patterns based on the order of presentation of the rubric performance levels?

Descriptive data for sample students' fixation times during the initial reading of the rubric, number of gaze transitions between each rubric PL (i.e., the combination of all the cells indicating the same PL) and the landscape during the oral analysis, and performance in the written analysis task for the whole sample and all experimental conditions can be found in Table 2. In addition, we identified outlier values ($\pm 3SD$) in data from fixation times on rubric cells and gaze transitions between each rubric PL and the picture of the

Table 2 Mean and *SD* of fixation time, PL4-picture transitions, and task performance for each experimental condition at step 1. Skewness and kurtosis for the whole dataset

Variable		Spanish speakers		Basque-Spanish speakers		Whole dataset	
		1–4 PL order Mean (*SD*)	4–1 PL order Mean (*SD*)	1–4 PL order Mean (*SD*)	4–1 PL order Mean (*SD*)	Skew	Kurt
Fixation time[1]						2.22	8.86
	Cell PL1	57.34 (29.00)	36.13 (32.41)	68.59 (29.81)	29.51 (17.61)		
	Cell PL2	42.49 (19.38)	34.19 (21.40)	49.55 (21.41)	22.69 (17.71)		
	Cell PL3	34.24 (18.59)	43.44 (28.11)	39.80 (22.25)	28.14 (15.59)		
	Cell PL4	46.59 (31.23)	91.82 (64.55)	62.50 (32.22)	48.36 (19.29)		
PL1-picture transitions in oral analysis		1.05 (1.61)	0.35 (0.94)	1.84 (3.30)	2.22 (4.63)	4.35	23.30
PL2-picture transitions in oral analysis		0.50 (0.89)	0.78 (1.11)	1.58 (2.61)	0.39 (0.78)	2.88	10.52
PL3-picture transitions in oral analysis		0.75 (1.02)	0.70 (1.11)	1.68 (2.16)	0.78 (1.11)	2.16	6.47
PL4-picture transitions in oral analysis		4.65 (5.32)	10.57 (10.17)	5.37 (5.30)	7.33 (9.77)	1.70	3.48
Task performance in written analysis		5.33 (1.45)	5.96 (1.16)	5.70 (1.08)	5.17 (1.24)	0.23	− 0.26

[1]In milliseconds per character. Skew: Skewness. Kurt: Kurtosis

landscape, which represented 1.56, 2.54, 2.50, 1.25, and 2.38% of the dataset, respectively. As reported above, these variables were thus log_{10}-transformed before performing the statistical analyses. Data from task performance in the written analysis of the landscape were approximately normally distributed (see Table 2) and no outlier value was found.

We also explored possible associations between eye-movement measures, students' university entry score, and scores in the written analysis task. A table including all these correlations is presented on Appendix H. Importantly, only fixation time on rubric PL4 -i.e., highest performance level- and the number of gaze transitions between rubric PL4 and the picture were significantly correlated (positively) to students' performance in the written analysis task ($r = 0.32$, $p < 0.01$ and $r = 0.22$, $p \leq 0.05$, respectively).

We first examined whether students' dominant language was moderated differences in students' rubric reading pattern during the initial reading of the rubric across rubric PL order conditions. The LMM for students' fixation time including the interaction term of cell PL, rubric PL order, and dominant language status as fixed factors showed differences across language groups in their reading patterns (an overview of the model's results can be seen in Table 3; pairwise comparisons in Table 4). Among all the possible Tukey-corrected pairwise comparisons, we focus on those comparing fixation times on PL4 vs. the other three rubric PL, as fixation time on PL4 was the only one positively associated with task performance in the written analysis of the landscape. As shown in Table 3, both language groups invested similar time on PL4, as compared to the other three rubric PL (all $ps > 0.57$). However, differences in students' reading pattern of the rubric emerged when examining the reading of the rubric with 4–1 PL order. Whereas the only significant difference in the BSS group appeared between fixation time on PL2 and on PL4, with increased reading time of PL4 ($t = 4.15$, $p < 0.01$), SDS students spent significantly more time on PL4 than on the other three rubric PL (vs PL1: $t = 5.92$, $p < 0.001$; vs PL2: $t = 4.98$, $p < 0.01$; vs PL3: $t = 3.94$, $p = 0.02$; see Table 4 and Fig. 2).

Thus, although there were no significant differences across language groups once post hoc comparisons were Tukey-corrected in fixation times on PL4 (see Table 4), our findings above suggest that SDS showed a more adaptive reading of the rubric than BSS. In other words, whereas the rubric with the 4–1 PL order fostered a reading pattern especially focused on rubric PL4 in the SDS group, this was not the case for the BSS students.

RQ2: How is students' dominant language associated with their task performance (written landscape analysis) when using rubrics?

When analyzing students' rubric reading behavior during the landscape oral analysis, we focused on gaze transitions between PL4 cells and the picture of the landscape, as fixation times revealed to be associated with better task performance in the written landscape analysis only for these cells. Results from a GLM including language group and rubric PL order as fixed factors showed that, although SDS did more transitions than BSS among those students who read the rubric with 4–1 PL order, the interaction between both factors was not significant (see Table 5). Thus, both language groups made similar number of transitions between PL4 and the landscape regardless of the rubric PL order (see also Fig. 3).

Lastly, we examined whether student performance in the landscape written analysis mirrored differences in rubric reading behavior across language groups and rubric PL order conditions. According to the differences in rubric reading patterns, the results from a GLM including the interaction term of dominant language status and rubric PL order as fixed factors revealed a significant interaction (estimate = 1.16, $SE = 0.56$, CI 95% [0.07, 2,25],

Table 3 LMM of the interaction effect of rubric PL order, cell PL, and dominant language group on fixation times

	Estimate	*SE*	CI 95%	*t*	*p*
Intercept	1.61	0.06	[1.49, 1.72]	27.85	< .001
Dom. language [a]	0.08	0.07	[−0.07, 0.22]	1.05	.30
PL order [b]	−0.20	0.07	[−0.34, −0.06]	−2.84	< .01
PL2 [c]	−0.07	0.06	[−0.19, 0.04]	−1.24	.22
PL3[c]	−0.13	0.06	[−0.25, −0.02]	−2.22	.03
PL4[c]	−0.10	0.06	[−0.21, 0.02]	−1.64	.11
PL2: PL order	0.13	0.06	[0.004, 0.25]	2.02	.04
PL3: PL order	0.25	0.06	[0.12, 0.38]	3.90	< .001
PL4: PL order	0.44	0.06	[0.32, 0.56]	6.96	< .001
Dom. language: PL order	−0.15	0.10	[−0.35, 0.96]	−1.42	.16
Dom. language: PL2	−0.06	0.06	[−0.18, 0.07]	−0.87	.38
Dom. language: PL3	−0.11	0.06	[−0.24, 0.02]	−1.69	.09
Dom. language: PL4	−0.04	0.06	[−0.17, 0.08]	−0.66	.51
Dom. language: PL2: PL order	−0.05	0.09	[−0.23, 0.14]	−0.52	.60
Dom. language: PL3: PL order	0.01	0.09	[−0.17, 0.19]	0.10	.92
Dom. language: PL4: PL order	−0.09	0.09	[−0.27, 0.09]	−0.95	.34
Random effects					
	Variance	*SD*		Correlation	
Participant (intercept)	0.03	0.17		-	
Cell (intercept)	0.004	0.06		-	
Model fit					
Marginal R^2	Conditional R^2				
.10	.34				

PL, rubric performance level; [a]reference value: SDS; [b]reference value: 1–4 PL order; [c]reference value: PL1

Table 4 Post hoc pairwise comparisons (Tukey corrected) between students' fixation times on each rubric PL among rubric order conditions

PL order	Comparison	Estimate	CI 95%	*SE*	*t*	*p*
1–4	SDS PL4 vs PL1	−0.10	[−0.31, 0.12]	0.06	−1.64	.95
	SDS PL4 vs PL2	−0.02	[−0.24, 0.19]	0.06	−0.40	1.00
	SDS PL4 vs PL3	0.04	[−0.18, 0.25]	0.06	0.59	1.00
	ESS PL4 vs PL1	−0.14	[−0.35, 0.07]	0.06	−2.38	.57
	ESS PL4 vs PL2	−0.01	[0.22, −0.20]	0.06	−0.18	1.00
	ESS PL4 vs PL3	0.10	[−0.11, 0.32]	0.06	1.74	.92
4–1	SDS PL4 vs PL1	0.34	[0.13, 0.55]	0.06	5.92	< .001
	SDS PL4 vs PL2	0.29	[0.08, 0.50]	0.06	4.98	.001
	SDS PL4 vs PL3	0.23	[0.02, 0.43]	0.06	3.94	.02
	ESS PL4 vs PL1	0.21	[−0.01, 0.43]	0.06	3.41	.08
	ESS PL4 vs PL2	0.26	[0.04, 0.41]	0.06	4.15	.01
	ESS PL4 vs PL3	0.20	[−0.02, 0.42]	0.06	3.18	.14
1–4	PL4: SDS vs BSS	−0.03	[−0.29, 0.22]	0.07	−0.46	1.00
4–1	PL4: SDS vs BSS	0.20	[−0,05, 0.45]	0.07	2.82	.27

PL, rubric performance level; *SDS*, Spanish-dominant speaker; *BSS*, Basque-Spanish speaker

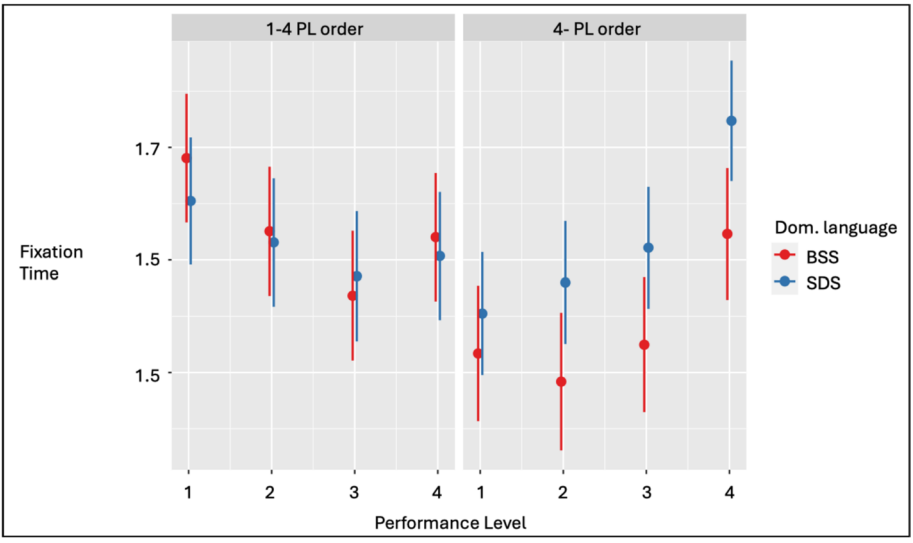

Fig. 2 Fixation time on each cell PL across rubric PL order conditions and dominant language status

$t = 2.09$, $p = 0.04$). As can be seen in Fig. 4, whereas there was no difference in task performance between student dominant language groups among those who read the rubric with the 1–4 PL order (estimate = 0.37, $SE = 0.40$, CI 95% [−0.40, 1.15], $t = 0.94$, $p = 0.35$), SDS outperformed BSS when reading the rubric with the 4–1 PL order (estimate = 0.79, $SE = 0.39$, CI 95% [0.03, 1.55], $t = 2.03$, $p = 0.046$). However, it is noteworthy that BSS students who read the rubric 1–4 PL order performed slightly better than SDS (but not

Table 5 GLM for PL4-picture gaze transitions including the interaction term between language group and rubric PL order

	Estimate	*SE*	CI 95%	*t*	*p*
Intercept	4.65	1.80	[1.05, 8.24]	2.58	.01
Dom. language (ESS)	0.72	2.58	[−4.43, 5.86]	.28	.78
PL order (4–1)	5.91	2.46	[1.00, 10.82]	2.40	.02
Dom. language (ESS): PL order (4–1)	−3.95	3.62	[−2.25, −0.07]	−1.09	.28

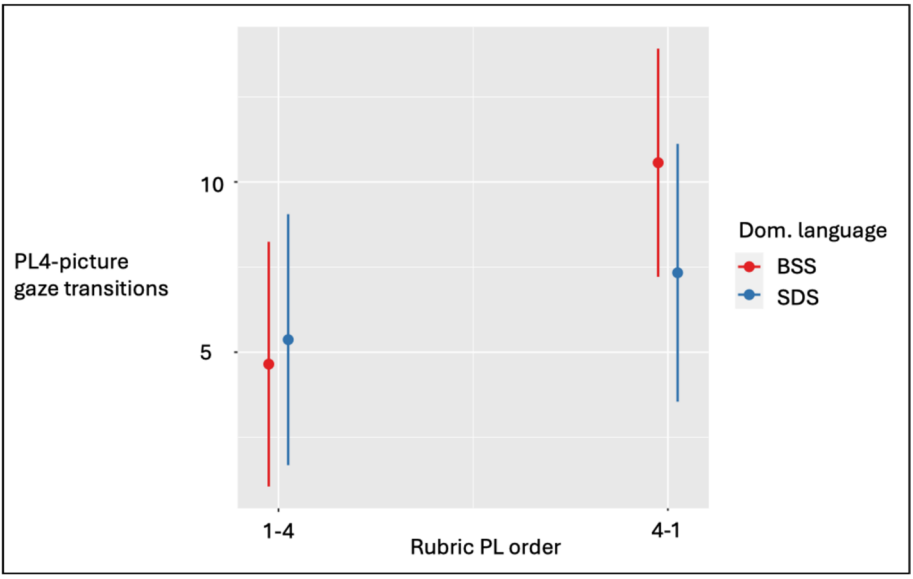

Fig. 3 PL4-picture gaze transitions during landscape oral analysis for each rubric PL order and language status

significant), so that their performance was similar to that of SDS who read the rubric with 4–1 PL order (see Fig. 4) (Table 6).

Discussion

Our aim was to investigate how the dominant language of students moderates students' rubric reading patterns across two different rubric order presentation and its association to task performance in a written landscape analysis. We manipulated the performance levels order of presentation (highest to lowest vs. lowest to highest). Importantly, our results revealed a significant impact of rubric presentation order on students reading patterns and a moderator effect of dominant language. This finding aligns with previous research on the critical role of well-designed interventions in reading and writing education (Graham & Hebert, 2010; Peltzer et al., 2024; Rijlaarsdam et al., 2024). More specifically we investigated two research questions and found the following.

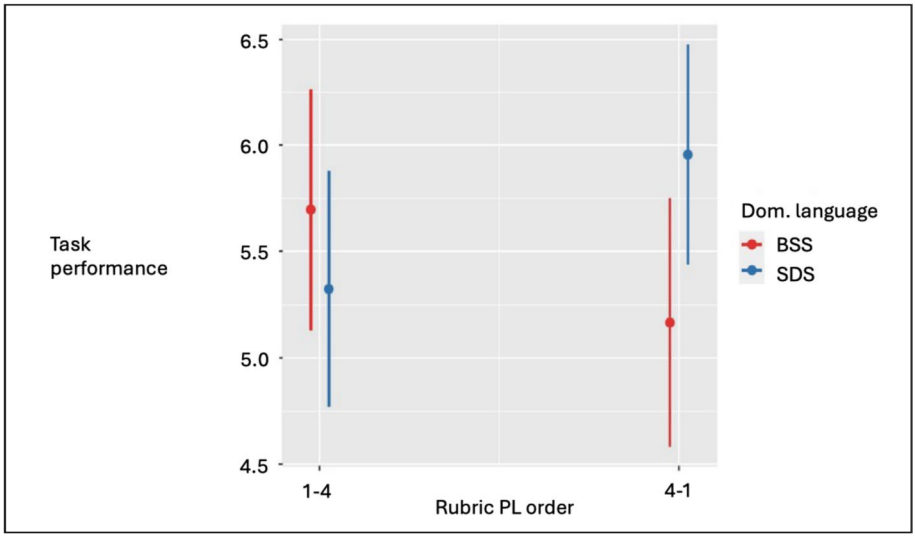

Fig. 4 Students' task performance scores on written analysis 1 for each rubric PL order and language status

Table 6 GLM for students' performance in landscape written analysis including the interaction term between language group and rubric PL order

	Estimate	*SE*	CI 95%	*t*	*p*
Intercept	5.32	0.28	[4.78, 5.87]	19.21	< .001
Dom. language (BSS)	0.37	0.40	[− 0.41, 1.15]	0.94	.35
PL order (4–1)	0.63	0.38	[− 0.11, 1.37]	1.67	.10
Dom. language (BSS): PL order (4–1)	− 1.16	0.56	[− 2.25, − 0.07]	− 2.09	.04

Reading patterns

Our results revealed significant differences in how Basque-Spanish speakers and Spanish-dominant speakers interacted with the rubric, particularly in the 4–1 performance level (PL) order condition (i.e., highest to lowest). Specifically, while both language groups showed similar fixation times on PL4 overall, the Spanish-dominant speakers group exhibited a more pronounced and adaptive reading pattern when the rubric was presented in the 4–1 PL order, spending significantly more time on PL4 compared to the other three PLs. This finding aligns with previous research suggesting that rubric structure and presentation order can impact how individuals interact with and interpret rubr ic criteria (Winke & Lim, 2015).

The differential reading patterns observed in our study are consistent with the theoretical underpinnings of feedback and formative assessment. According to Hattie and Timperley (2007), effective feedback must be clear and targeted, addressing specific areas that need improvement. Our results suggest that the structured feedback provided by the rubric, particularly in the 4–1 PL order, is linked to more effective focus on the most advanced performance level (PL4) for Spanish-dominant students than for Basque Spanish students.

These findings underscore the importance of considering linguistic diversity when designing rubrics, as the dominant language appears to shape how students process and utilize feedback. Moreover, the observed differences is supported, by previous work that emphasized that rubrics can enhance self-regulated learning by clarifying expectations and guiding students in identifying areas for improvement (e.g., Brookhart, 2018; Panadero & Jonsson, 2013).

The observed pattern could also be explained by the inherent readability or perceived logical flow of the 4–1 order, particularly for Spanish-dominant students who may find this structure more intuitive. Research on attentional patterns (e.g., Sweller et al., 2011; Narciss, 2008) suggests that learners often focus more on the first information they encounter, a "primacy effect" that could apply here regardless of whether the initial PL is the highest or the lowest.

In addition, the impact of linguistic background cannot be discounted. Spanish-dominant speakers may process the 4–1 order more effectively due to a greater alignment with their cognitive and linguistic frameworks. Conversely, Basque Spanish speakers may experience a higher cognitive load when engaging with rubrics in Spanish, which could diminish their ability to benefit from the order manipulation. These findings highlight the interaction between rubric design, cognitive processing, and linguistic background, emphasizing the need for further research to explore how linguistic and cognitive factors shape rubrics reading pattern and their latter association with task performance.

Additionally, our study contributes to the limited body of research employing eye-tracking technology to analyze students' interactions with rubrics. Previous studies, such as those by Winke & Lim (2015), focused primarily on professional raters, leaving a gap in understanding how students use rubrics. By examining eye movements, we provide empirical evidence on how linguistic factors affect rubric reading behavior. This methodological approach aligns with Panadero (2023) call for more process-oriented research in educational assessment, offering deeper insights into the cognitive processes underlying students' use of rubrics. Our findings suggest that the interaction between dominant language and rubric design warrants further investigation, particularly to develop more inclusive and effective assessment tools that cater to the diverse linguistic backgrounds of students.

Task performance in written landscape analysis

The results indicated that the Spanish-dominant students outperformed the Basque Spanish students in their written landscape analysis performance when the rubric was presented in the 4–1 PL order. This finding suggests that the order in which rubric criteria are presented can significantly impact student performance, at least in written analysis task, particularly for those whose dominant language matches the task language. These results are in line with previous research that highlights the importance of rubric structure in guiding student performance (Andrade et al., 2010; Author, 2013). The observed differences in performance might be attributed to the clarity and focus provided by the 4–1 PL order, which potentially made it easier for Spanish-dominant students to align their writing with the highest performance criteria (PL4).

The significant interaction between dominant language status and rubric PL order suggests that the cognitive processes involved in interpreting and applying rubric criteria can vary based on linguistic background. This supports the theoretical framework proposed by Hattie and Timperley (2007), which emphasizes the role of feedback in reducing

discrepancies between current performance and desired goals. In this context, the 4–1 PL order rubric appears to provide clearer guidance, particularly benefiting Spanish-dominant students. This finding also aligns with Narciss's (2008) Interactive Tutoring Feedback Model, which advocates for feedback that is specific and detailed, aiding learners in adjusting their cognitive processing. The performance improvement seen in Spanish-dominant students with the 4–1 PL order rubric highlights the importance of designing rubrics that cater to the cognitive and linguistic needs of diverse student populations.

Additionally, our study contributes to the body of literature on the impact of rubric design on student performance, emphasizing the role of linguistic factors. Previous studies have primarily focused on the general effectiveness of rubrics (Brookhart & Chen, 2015), but our findings underscore the need to consider how dominant language influences the interaction with rubrics. The fact that Basque Spanish students performed better with the 1–4 PL order, though not significantly, suggests that different rubric structures might be effective depending on the students' linguistic background. This aligns with the work of Panadero and Lipnevich (2022), who emphasized the importance of feedback message and student characteristics in feedback effectiveness. Our results suggest that further research is needed to explore how rubric design can be optimized for multilingual settings, ensuring that all students, regardless of their dominant language, can benefit equally from rubric-based feedback.

Finally, our findings, which show the significant impact of rubric presentation order and dominant language on student performance, align with comprehensive approaches to writing interventions (Graham & Hebert, 2010). For example, Rijlaarsdam and colleagues (Rijlaarsdam et al., 2017) highlighted the importance of structured instructional design and the role of feedback in enhancing writing skills. This supports our conclusion that strategic rubric design is essential for optimizing formative assessment practices and improving student outcomes.

Limitations

Firstly, our study includes limitations related to the specific characteristics of the participants and the method of recruitment. Participants were undergraduate students recruited through convenience sampling. While the sample included students from six different undergraduate programs, enhancing its diversity, we acknowledge that convenience sampling inherently limits the representativeness of the sample. Out of 174 students contacted, 138 participated in the study, and only 80 were included after data quality control. This sample size is substantial for an eye-tracking study, but it does not allow for full generalizability of the findings. Future studies should aim to include larger and more representative samples to increase the generalizability of results.

Secondly, our final sample size was reduced to 80 participants due to the rigorous quality control applied to the eye-tracking data. Specifically, participants with data loss exceeding 25% or calibration errors beyond 0.5° accuracy were excluded to ensure data reliability (Holmqvist et al., 2011). There are no indications in the consulted literature about an acceptable proportion of data loss per participant. However, we consider that 75% of the data is large enough to get substantial information about each participant's performance. Importantly, these exclusions were random, unrelated to participant characteristics, and do not appear to have introduced systematic bias into the sample. While this reduction aligns with standard practices in eye-tracking studies, it may have implications for the generalizability of our findings. Nonetheless, we acknowledge that the reduced sample size should

be considered when interpreting the results, and future studies could aim to increase the initial sample size to account for similar data quality constraints.

Thirdly, dominant language was determined based on students' self-perceptions of their linguistic background rather than objective tests. We acknowledge that this approach introduces limitations, particularly in the classification of Spanish-dominant Speakers and Basque-Spanish Speakers (BSS). Future research could use standardized proficiency tests to complement self-reported measures and provide a more nuanced understanding of language dominance.

Fourthly, the rubric was written exclusively in Spanish, and we did not counterbalance the rubric language with a Basque version, nor did we explore the rubric language as an experimental factor. This choice limits the conclusions we can draw about the role of dominant language. For instance, it is unclear whether the observed effects are due to language dominance, familiarity with the rubric's language, or other factors. Future studies should explore the influence of rubric language by including rubrics in both Spanish and Basque to provide a more comprehensive understanding of these effects.

Lastly, our statistical analyses may be limited by issues of statistical power. Specifically, for the analysis of fixation times using LMM, there was no comparable previous study to facilitate an a priori calculation of statistical power. According to Kumle et al. (2021), raw data from prior studies that explored similar variables are necessary for calculating a priori statistical power for LMM analyses. Additionally, our a priori power analysis for general linear model (GLM) analyses indicated insufficient power to detect even a medium-sized interaction effect within our 2×2 between-participant experimental design. Consequently, any null effects observed should be interpreted with caution, and definitive conclusions cannot be confidently drawn from them. Despite these limitations, it is noteworthy that our sample size is relatively large for an eye-tracking study. In particular, the only other known study that explored eye-tracking data in the context of rubric use included 11 participants (Winke & Lim, 2015).

These limitations underscore the exploratory nature of our findings, particularly regarding the relationship between dominant language and rubric use. We caution against overgeneralizing our results and advocate for further research to address these limitations systematically. Although not a limitation per se, it is important to note that this research project involved a comprehensive multimodal data collection approach, encompassing self-reports, think-aloud protocols, electrodermal activity, task performance, and eye-tracking. However, the focus of this study was primarily on the eye-tracking data.

Conclusion

Our research underscores the importance of considering dominant language in the design and implementation of rubrics. The adaptive reading patterns exhibited by SDS students in the 4–1 PL order condition highlight how rubric design can enhance student interaction with the rubric (i.e., reading) and performance (i.e., written landscape analysis), particularly in linguistically diverse educational settings. Our findings suggest that rubric structure design is associated with variations in student performance, depending on their dominant languages. This has profound implications for educators and researchers aiming to optimize formative assessment practices and develop more effective, inclusive educational tools. By tailoring feedback mechanisms and rubric designs to accommodate the diverse linguistic backgrounds of students, we can better support all learners in achieving their academic goals and foster a more equitable learning environment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10212-025-00951-9.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. (1) Basque Government Call for Grants to support the activities of research groups of the Basque University System (2022–2025) project reference IT1624-22. (2) Basque Country Equipment 2021 call. Project: Eye tracker. Reference: EC21_2021_1_0004. (3) Ayudas a los Agentes del Sistema Andaluz del Conocimiento para la Contratación de Personal Investigador Doctor (PAIDI DOCTOR 21; Regional Government of Andalusia, Spain) to the second author.

Declarations

Ethics approval This research has been approved by the ethics committee from Comité de ética de la investigación, Universidad de Deusto. Reference: ETK-5/21–22. PI: Ernesto Panadero.

Conflict of interest The authors declare no competing interests.

References

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159–181. https://doi.org/10.1080/02602930600801928

Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice, 17*(2), 199–214. https://doi.org/10.1080/09695941003696172

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1–11. Retrieved from http://pareonline.net/getvn.asp?v=10&n=3

Aronin, L. (2016). Dominant language constellations: A new perspective on multilingualism. *Journal of Multilingual and Multicultural Development, 44*(5), 541–542. https://doi.org/10.1080/01434632.2021.1993550

Barto, K. (2023). *MuMIn: Multi-model inference. R package version 1.47.5*. https://cran.r-project.org/web/packages/MuMIn/index.html

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2021). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–27.1*. https://cran.r-project.org/web/packages/lme4/index.html

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7*(3), 325–339. https://doi.org/10.1111/j.1467-7687.2004.00351.x

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education, 3*, 22. https://doi.org/10.3389/feduc.2018.00022

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review, 67*(3), 343–368. https://doi.org/10.1080/00131911.2014.929565

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science, 11*(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Cenoz, J. (2009). *Towards multilingual education: Basque educational research from an international perspective*. Multilingual Matters.

Conklin, K., Pellicer-Sánchez, A., spsampsps Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research.* Cambridge University Press.

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation In Higher Education*, 1–14. https://doi.org/10.1080/02602938.2015.1111294

Eustat (2021). *Población de la C.A. de Euskadi por ámbitos territoriales, primera lengua y lengua hablada en casa*. https://www.eustat.eus/elementos/ele0021900/poblacion-de-la-ca-de-euskadi-por-ambitos-territoriales-primera-lengua-y-lengua-hablada-en-casa/tbl0021956_c.html

Graham, S., & Hebert, M. (2010). Writing to read: Evidence for how writing can improve reading. Carnegie Corporation of New York. https://lincs.ed.gov/publications/pdf/writing_to_read.pdf

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* OUP Oxford.

Humphry, S. M., & Heldsinger, S. A. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher, 43*(5), 253–263. https://doi.org/10.3102/0013189x14542154

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Lasekan, O. (2020). Using the interpersonal relationship communication approach to assess language dominance in both written and oral communication among multilinguals. *Journal Title Not Provided, 8*, 164–176. https://doi.org/10.20896/saci.v8i1.623

Leibowitz, B. (2005). Learning in an additional language in a multilingual society: The sociopolitical context in South Africa. *Per Linguam, 21*(2), 3–16.

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M, Jun, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *emmeans: Estimated marginal means for R*. https://cran.r-project.org/web/packages/emmeans/index.html

Lipnevich, A. A., Panadero, E., & Calistro, T. (2022). Unraveling the effects of rubrics and exemplars on student writing performance. *Journal of Experimental Psychology: Applied, 29*, 136. https://doi.org/10.1037/xap0000434

Lowinger, R. J., He, Z., Lin, M., & Chang, M. (2014). The impact of academic self-efficacy, acculturation difficulties, and language abilities on procrastination behavior in Chinese international students. *College Student Journal, 48*(1), 141–152.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition, 11*(01), 81–93. https://doi.org/10.1017/S1366728907003227

McDonald, R. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125–144). Lawrence Erlbaum.

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806–813. https://doi.org/10.1016/j.lindif.2012.04.007

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551–576. https://doi.org/10.14204/ejrep.30.12200

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*(0), 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review, 30*,. https://doi.org/10.1016/j.edurev.2020.100329

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback typologies and models: Towards an integrative model of feedback elements. *Educational Research Review*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Panadero, E. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. *Educational Psychologist, 58*(3), 193–204. https://doi.org/10.1080/00461520.2023.2223642

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: a Meta-analytic Review. *Educational Psychology Review, 35*(4), 113. https://doi.org/10.1007/s10648-023-09823-4

Peltzer, K., Lira Lorca, A., Krause, U. M., & Busse, V. (2024). Effects of formative feedback on argumentative writing in English and cross-linguistic transfer to German. *Learning and Instruction, 92*, 101935. https://doi.org/10.1016/j.learninstruc.2024.101935

Poeste, M., Müller, N., & Arnaus Gil, L. (2019). Code-mixing and language dominance: Bilingual, trilingual and multilingual children compared. *International Journal of Multilingualism, 16*(4), 459–491. https://doi.org/10.1080/14790718.2019.1569017

Postmes, L., Bouwmeester, R., de Kleijn, R., & van der Schaaf, M. F. (2023). Supervisors' untrained postgraduate rubric use for formative and summative purposes. *Assessment & Evaluation in Higher Education, 48*(1), 41–55. https://doi.org/10.1080/02602938.2021.2021390

R Core Team. (2023). *The R stats package version 4.4.0.* https://www.R-project.org/

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448. https://doi.org/10.1080/02602930902862859

Rijlaarsdam, G., Janssen, T., van Weijen, D., & van den Bergh, H. (2017). Improving writing in primary schools through a comprehensive writing program. *The Journal of Writing Research, 9*(2), 173–225. https://doi.org/10.17239/jowr-2017.09.02.04

Rijlaarsdam, G., van Ockenburg, L., Mateos, M., & van Weijen, D. (2024). Essences for designing interventions in writing research. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (3rd ed.). The Guilford Press.

Rivero-Contreras, M., Engelhardt, P. E., & Saldana, D. (2023). Do easy-to-read adaptations really facilitate sentence processing for adults with a lower level of education? An experimental eye-tracking study. *Learning and Instruction, 84*, 101731. https://doi.org/10.1016/j.learninstruc.2022.101731

Schuth, E., Köhne, J., & Weinert, S. (2017). Influence of vocabulary knowledge on school performance. *Journal of Educational Research, 110*(6), 559–573.

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J.,... Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, spsampsps R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). Cambridge University Press. https://doi.org/10.1017/CBO9780511844744.004

Tobiipro. (2021). Pro Lab User Manual v.1.181.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education, 14*(3), 281–294. https://doi.org/10.1080/09695940701591867

Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics, 5*, 375–393. https://doi.org/10.1146/annurev-linguistics-011718-011701

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7*, 769. https://doi.org/10.3389/fpsyg.2016.00769

Turley, E. D., & Gallagher, C. W. (2008). On the "uses" of rubrics: Reframing the great rubric debate. *The English Journal, 97*(4), 87–92. https://doi.org/10.2307/30047253

Winke, P., & Lim, H. (2015). ESL essay raters' cognitive processes in applying the Jacobs et al. rubric: An eye-movement study. *Assessing Writing, 25*, 38–54. https://doi.org/10.1016/j.asw.2015.05.002

Yan, Z. (2018). Student self-assessment practices: The role of gender, school level and goal orientation. *Assessment in Education: Principles, Policy & Practice, 25*(2), 183–199. https://doi.org/10.1080/0969594X.2016.1218324

Yan, Z., Brown, G. T. L., Lee, J. C. K., & Qiu, X. L. (2020). Student self-assessment: Why do they do it? *Educational Psychology, 40*(4), 509–532. https://doi.org/10.1080/01443410.2019.1672038

Zhuang, Y. (2023). Boosting self-motivation in English language learning: The role of innovative teaching strategies. *Journal of Language Teaching and Research, 24*(2), 234–249.

Authors' "Current Themes of Research" and "Most Relevant Publications" in the field of Psychology of Education.

Ernesto Panadero. His research focuses on self-regulated learning and educational assessment. Some of his most relevant publications are:.

Panadero, E. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. Educational Psychologist, 58(3), 193-204. https://doi.org/10.1080/00461520.2023.2223642.

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: a Meta-analytic Review. Educational Psychology Review, 35(4), 113. https://doi.org/10.1007/s10648-023-09823-4.

Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: Intrapersonal and interpersonal factors. Assessment & Evaluation In Higher Education, 48(8), 1053-1075. https://doi.org/10.1080/02602938.2023.2164884.

Pablo Delgado. His research focuses on reading on digital media, with a special interest in the differences between reading in print and on screen. He also investigates differences between learning from texts and learning from videos in digital contexts, and sourcing (i.e., paying attention to and evaluating trustworthiness of the sources of information) when reading on the Internet.

Some of his most relevant publications are:.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review*, *25*, 23-38. https://doi.org/10.1016/j.edurev.2018.09.003.

Delgado, P., & Salmerón, L. (2021). The inattentive on-screen reading: Reading medium affects attention and reading comprehension under time pressure. *Learning and Instruction*, *71*, Article 101396. https://doi.org/10.1016/j.learninstruc.2020.101396.

Delgado, P., Anmarkrud, Ø., Avila, V., Altamura, L., Chireac, S. M., Pérez, A., & Salmerón, L. (2022). Learning from text and video blogs: comprehension effects on secondary school students. *Education and Information Technologies*, *27*, 4249-5275. https://doi.org/10.1007/s10639-021-10819-2.

Delgado, P., Stang Lund, E., Salmerón, L., & Bråten, I. (2020). To click or not to click: Investigating conflict detection and sourcing in a multiple document hypertext environment. *Reading and Writing*, *33*, 2049-2072. https://doi.org/10.1007/s11145-020-10030-8.

Salmerón, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*, *36*, 1038-1051.https://doi.org/10.1111/jcal.12458.

Lucía Barrenetxea-Mínguez. Her research focuses on gifted education and teacher training. Some of her most relevant publications are:.

Barrenetxea-Mínguez, L., Galindo-Domínguez, H., & Yániz, C. (2024). Las Actitudes educativas del profesorado hacia las altas capacidades: Validación de la escala AEPAC. Profesorado, Revista de Currículum y Formación del Profesorado, 28(1), 101-124. https://doi.org/10.30827/profesorado.v28i1.29541.

Díaz-Iso, A., Gutiérrez-Fernández, N., & Barrenetxea-Mínguez, L. (2023). Conectando teoría y práctica: la metodología de Aprendizaje-Servicio en la didáctica de las lenguas en Educación Superior. Aula Abierta, 52(4), 343–350. https://doi.org/10.17811/rifie.52.4.2023.343-350.

Barrenetxea-Mínguez, L. B., & Izaguirre, M. M. (2020). Relevancia de la formación docente para la inclusión educativa del alumnado con altas capacidades intelectuales. Atenas, 1(49), 1-19.

David Zamorano. His research focuses on formative assessment, specifically peer feedback. Some of his most relevant publications are:.

Ocampo, J. C., Panadero, E., Zamorano, D., Sánchez-Iglesias, I., & Diez Ruiz, F. (2023). The effects of gender and training on peer feedback characteristics. Assessment & Evaluation in Higher Education, 1-17. https://doi.org/10.1080/02602938.2023.2286432.

Fraile, J., Gil-Izquierdo, M., Zamorano, D., & Sánchez-Iglesias, I. (2020). Self-regulated learning and formative assessment process on group work. Revista Electrónica de Investigación y Evaluación Educativa (RELIEVE), 26(1), M5. https://doi.org/10.7203/relieve.26.1.17402.

Leire Pinedo. Her research focuses on self-assessment, examining how feedback mechanisms impact these processes. Some of her most relevant publications are:.

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: a Meta-analytic Review. Educational Psychology Review, 35(4), 113. https://doi.org/10.1007/s10648-023-09823-4.

Pinedo, L., Panadero, E., Fernández-Ruiz, J., Rodríguez-Hernández, C. (2023). Students' experiences in self-assessment: training, processes and feedback use in secondary and higher education, Assessment in Education: Principles, Policy & Practice, 30:5-6, 448-480, https://doi.org/10.1080/0969594X.2023.2284630.

Alazne Fernández Ortube. Her research focuses on self-regulated learning and rubrics. Her most relevant publications are:.

Panadero, E., Fernández Ortube, A., Krebs, R., & Roelle, J. (2024). Analysis of online rubric platforms: advancing toward erubrics. Assessment & Evaluation in Higher Education, 1–19. https://doi.org/10.1080/02602938.2024.2345657.

Authors and Affiliations

Ernesto Panadero[1,2,3] · Pablo Delgado[4] · Lucía Barrenetxea-Mínguez[2] · David Zamorano[2] · Leire Pinedo[2] · Alazne Fernández-Ortube[2]

✉ Lucía Barrenetxea-Mínguez
luciabarrenetxea@deusto.es

1 Centre for Assessment Research Policy and Practice in Education, School of Policy & Practice, Institute of Education, Dublin City University, Dublin, Ireland

2 Education Regulated Learning and Assessment (ERLA Group), Facultad de Educación y Deporte, Universidad de Deusto, 227 Office, 48007 Bilbao, Spain

3 IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

4 Laboratorio de Diversidad, Cognición y Lenguaje, Universidad de Sevilla, Seville, Spain