

# LINGUISTIC SECURITY IN ONLINE LEARNING: IMPROVING ASSESSMENT DESIGN FOR THE WRITING PROCESS

Aponte-Alequín HA \*

*Division of Institutional Research and Assessment, University of Puerto Rico Río Piedras Campus, Puerto Rico,  
USA*

---

**Abstract:** This study examines how revising assessment language can foster linguistic security in a first-year online Spanish Basic Writing course. Student perceptions of two short-answer items that had repeatedly underperformed were analyzed, asking whether the prompts' wording –rather than their linguistic content– was provoking confusion or discomfort. A post-test questionnaire captured immediate reactions to each item's clarity and perceived relevance, along with self-reported reasons for dissatisfaction. Findings show unanimous dissatisfaction with both prompts and distinct patterns across reasons, pointing to clarity issues in a syntax item and discomfort related to stigmatizing terminology in a speech-levels task. Guided by these patterns, revisions were introduced that neutralize affective content, align prompts with the taught scope, and make the requested operation explicit, such as naming the syntactic category or identifying the register mismatch. The scoring guide was also recalibrated to separate mastery of linguistic content from success at interpreting the prompt, and a brief communication was prepared for students explaining the changes to promote transparency and trust. The approach illustrates how a concise, course-embedded perception check can reveal construct-irrelevant barriers and guide targeted wording changes that protect validity. The process is practical for instructors who manage online basic-writing assessments and can be replicated in similar courses seeking to strengthen linguistic security while maintaining cognitive demand.

**Keywords:** basic writing, online learning, student learning assessment, applied linguistics, linguistic insecurity, assessment design

---

## Introduction

Linguistic security and linguistic clarity are central to fair assessment in online basic-writing contexts. In the first-year Spanish Basic Writing Workshop (ESPA 3003) at the University of Puerto Rico, Río Piedras Campus, instructors observed that several prompts in the online test produced unexpected difficulty not aligned with the intended learning outcomes. This pattern raised a validity concern: wording, discourse framing, or affective load in the prompts –rather than cognitive demand– might be introducing construct-irrelevant variance. To examine that interface between language and validity, we administered a brief, post-test questionnaire immediately after the assessment, capturing students' perceptions of clarity, alignment with taught scope, and affective reactions for two underperforming items.

\*Corresponding Author's Email: [hector.aponte2@upr.edu](mailto:hector.aponte2@upr.edu)

The instructional setting provides a precise context and a useful stress test. Between 2019 and 2021, sections of ESPA 3003 taken by students in Communication fields (Journalism, Audiovisual Communication, Public Relations and Advertising) served as a pilot for strengthening learning-assessment practices. Across three semesters (cohorts of approximately 110-120 students), two short-answer tasks consistently fell well below the 70% success threshold normally expected for passing, even as the same tasks reached expected rates in comparable face-to-face sections. The contrast suggests that the difficulty is not inherent to the linguistic content but is instead associated with the way the prompts are worded and framed, potentially undermining students' sense of linguistic security.

In this proceedings contribution, we therefore reframe the study around revising/improving assessment design rather than describing broad “transformations.” Our aims are twofold: (a) document the extent and profile of student dissatisfaction with the two items, and (b) propose targeted revisions that increase linguistic neutrality, tighten scope alignment with the course, and make the requested operation explicit (e.g., naming the relevant syntactic category or signaling a register mismatch). The questionnaire functions as a concise, course-embedded perception check that helps identify wording-induced barriers without lowering cognitive demand.

While the investigation centers on an online, Spanish-medium basic-writing course, the approach is designed to be portable. The protocol –pairing item-level perception data with small, evidence-based wording changes– can be replicated in other language courses and in non-language disciplines that rely on short prompts to elicit analytic responses. Subsequent sections outline the theoretical framework and methodology, report the results with visual summaries, and discuss practical revisions that support fairness and student trust in online assessment.

This study positions linguistic security and linguistic neutrality as the core constructs linking assessment language to validity in online basic-writing contexts. We treat linguistic security as students' perception that assessment prompts are clear, aligned to what was taught, and free of affective or stigmatizing triggers that might interfere with task focus. In contrast, linguistic insecurity refers to attitudinal dispositions that lead speakers to experience negative feelings about their language variety or particular features of it, thereby eroding confidence during performance (Trudgill & Hernández Campoy, 2007, pp. 182–183; Herrera & Díaz, 2021). Linguistic security, understood as the absence of such negative emotions toward the mother tongue in assessment settings (García & Fumero, 2010, p. 101), is not merely an affective ideal; it is a validity requirement whenever prompt wording risks drawing attention away from the intended construct.

### **Language- validity interface**

The interface between wording and validity is well documented in applied linguistics and assessment studies. Research on native-language pedagogy emphasizes that textual formulation, register cues, and discourse framing shape comprehension and performance in measurable ways (Cassany, Luna, & Sanz, 2007; García & Fumero, 2010). In testing terms, ambiguous task verbs, opaque error labels, or emotionally charged scenarios can introduce construct-irrelevant variance that suppresses performance independently of the targeted ability (Volkwein, 2010, 2011; Suskie, 2018). Empirical work further shows that sensitive content or stigmatizing terminology can activate discomfort and attentional shifts detrimental to metalinguistic tasks, especially when students must identify or name an error type (Fumero, 2021; Fumero, 2021, p. 120; Ogange et al., 2018). Our operational focus on revising/improving assessment design follows this literature by privileging small, testable wording adjustments over wholesale redesigns.

### **Online learning as a stress test for clarity**

The online setting heightens the need for precise, neutral language. Distinctions between well-designed online learning and emergency remote teaching underscore how modality constraints alter communication channels, feedback timing, and students' interpretive resources (Hodges et al., 2020; Meyer & Murrell, 2021). Without in-room clarification or prosodic cues, students rely more heavily on the literal wording of prompts. Under these conditions, micro-features—task verbs, scope markers, register labels—can disproportionately affect performance, which strengthens the case for systematic, evidence-based prompt revision.

### **Indirect assessment and student-centered evidence**

We integrate indirect assessment as a student-centered diagnostic to detect wording-induced barriers. Student perception data, when gathered with brief, targeted questionnaires, contribute credible evidence about instrument functioning and help close feedback loops in continuous improvement cycles (Jankowski et al., 2018; Ibarra-Sáiz et al., 2023; Suskie, 2018). In online contexts, indirect measures partially substitute for the lost immediacy of face-to-face observation (Kumar & Rani, 2021), enabling instructors to pinpoint where clarity, scope alignment, or affective load may be compromising validity. Our use of a short post-test questionnaire aligns with this stream: it elicits item-specific clarity/relevance judgments and reasons for dissatisfaction to guide precise edits rather than impressionistic changes.

## **Revising/improving assessment design: clarity, neutrality, and scoring**

A revising/improving stance prioritizes three levers:

- Clarity of the requested operation- Explicitly naming the linguistic category or register operation reduces ambiguity and stabilizes interpretations (Cassany et al., 2007; García & Fumero, 2010).
- Neutrality of content and framing- Removing stigmatizing or affect-laden elements lowers affective load and keeps attention on the construct (Fumero, 2021; Ogange et al., 2018).
- Rubric calibration- Scoring guides should separate construct mastery (e.g., syntactic knowledge) from success at decoding prompt language; this protects fairness when wording has been a barrier (Mace & Pearl, 2020).

Within an institutional assessment ecosystem, these levers fit standard cycles of evidence gathering and improvement (Volkwein, 2010, 2011; Jankowski et al., 2018), and they are readily portable to other disciplines that depend on brief analytic prompts.

## **Working propositions**

Guided by the foregoing, we advance two propositions that structure the study:

P1- Wording sensitivity: In online basic-writing tests, items that contain opaque task verbs or stigmatizing terminology will elicit higher dissatisfaction and lower success independent of construct difficulty (Fumero, 2021; García & Fumero, 2010; Ogange et al., 2018).

P2- Improvement through targeted revision: Brief, student-centered perception checks will identify specific wording barriers whose revision—via clearer task cues, neutralized content, and rubric calibration—should enhance linguistic security and protect validity without reducing cognitive demand (Ibarra-Sáiz et al., 2023; Suskie, 2018; Kumar & Rani, 2021).

In sum, this framework reframes the project as a practical exercise in revising/improving assessment design under online constraints. It aligns with institutional commitments to credible, student-centered evidence and offers a portable method that other programs can adopt to mitigate construct-irrelevant variance while maintaining rigor (Hodges et al., 2020; Meyer & Murrell, 2021; Jankowski et al., 2018; Volkwein, 2010, 2011).

## **Materials and Methods**

This study adopts a quantitative institutional-research design (Volkwein, 2011) to diagnose wording-related barriers in an online basic-writing test and to generate precise edits that revise/improve assessment design without lowering cognitive demand. The methodology operationalizes the constructs defined in Section 2 –linguistic security and linguistic neutrality–and privileges portability to comparable courses and programs.

### **Design and setting**

We implemented a descriptive, course-embedded protocol in ESPA 3003 (Spanish Basic Writing Workshop) at a large public research university. Instruction and testing occurred fully online through the institutional LMS (Moodle). The focal test comprised 50 short-answer items requiring students to identify linguistic errors or define language concepts. Across three prior semesters, the test showed acceptable internal consistency (Cronbach's  $\alpha = .751$ ). Following institutional guidelines for assessment cycles (Volkwein, 2011; Suskie, 2018), the present iteration added an indirect assessment component immediately after the test to capture students' item-level perceptions (Ibarra-Sáiz et al., 2023).

### **Participants and sampling**

Using non-probability convenience sampling typical of institutional improvement studies (Volkwein, 2011), we analyzed the full enrollment of  $n = 117$  first-year students in Communication majors (Journalism; Audiovisual Communication; Public Relations and Advertising) during the first semester of 2021–2022. Students were distributed across five online sections: 001 ( $n = 23$ ), 002 ( $n = 21$ ), 003 ( $n = 25$ ), 004 ( $n = 23$ ), and 005 ( $n = 25$ ). Given the course's introductory level, Spanish proficiency was relatively homogeneous, supporting meaningful comparisons.

### **Focal items and rationale**

Institutional trend analyses flagged two short-answer items with persistently low success ( $<10\%$  correct across earlier cohorts) despite explicit coverage in class, against an expected threshold of  $\sim 70\%$  for passing:

Item 1 (Syntax). Identify a specific syntactic fault in a sentence. The original scenario included violent content.

Item 2 (Speech levels). Provide an example of a stigmatized register ("vulgar speech") and justify its exclusion from formal expression.

Consistent with our focus on linguistic security, both items were excluded from course grading for the present administration to avoid penalizing students for wording-induced barriers (Trudgill & Hernández Campoy, 2007; Fumero, 2021).

### **Indirect assessment instrument**

Immediately after submitting the test, students completed a brief post-test questionnaire delivered in the LMS. For each focal item, the instrument first asked a binary satisfaction question (satisfied / dissatisfied), then—contingent on dissatisfaction—prompted a single-choice reason. Response structure followed recommendations to minimize cognitive load and maximize interpretability in online settings (Meyer & Muller, 2021; Meyer & Muller, 2021). Reasons were aligned with the constructs in Section 2:

1. Content distracts (affective load / non-neutral framing),
2. I don't understand what error to report (clarity of the requested operation),
3. This is outside our class scope (relevance / alignment), and
4. This prompt makes me uncomfortable (security / stigma activation).

A brief preface reminded students that the aim was to improve prompt wording, not to evaluate their ability, reinforcing a student-centered stance (Ibarra-Sáiz et al., 2023; Suskie, 2018).

### **Reliability and quality checks**

The long-form test's prior  $\alpha = .751$  provided baseline consistency. For the questionnaire used here, scale-level intraclass correlation coefficients  $> .90$  indicated high stability for the binary satisfaction prompt paired with the reason checklist, supporting its use to guide targeted revisions (Suskie, 2018; Volkwein, 2011).

### **Procedures**

The sequence was: (1) online instruction; (2) administration of the 50-item short-answer test; (3) immediate administration of the questionnaire for the two focal items; (4) export of anonymized, section-level data. The deferral of any clarifying discussion until after data capture preserved the naturalistic reading experience typical of online testing (Meyer & Muller, 2021). To protect students and preserve validity, the two focal items did not contribute to grades (Fumero, 2021; Trudgill & Hernández Campoy, 2007).

## **Measures and analysis**

We computed percent dissatisfied by item and distribution of reasons among dissatisfied respondents. Given the diagnostic purpose of the indirect instrument, descriptive statistics (percentages) were prioritized to answer the practical design questions at hand (Fumero, 2021; Suskie, 2018). The resulting profiles were then mapped to three revising/improving levers derived from Section 2—clarity of the requested operation, neutrality of content/framing, and rubric calibration—to formulate precise wording edits.

## **Ethics and portability**

Data were aggregated, non-identifiable course-level evidence gathered as part of routine assessment quality improvement; under institutional policy, no human-subjects review was required. The procedure is intentionally portable: any course using brief analytic prompts can embed a two-step perception check post-test to detect wording-induced variance and to guide small, evidence-based revisions within standard assessment cycles (Volkwein, 2011; Ibarra-Sáiz et al., 2023).

## **Results and Discussion**

Of the 117 students enrolled, 100 completed the post-test questionnaire. As shown in Table 1, satisfaction with the two focal prompts was 0%, with 100% of respondents marking “I am dissatisfied” for Item 1 (Syntax) and Item 2 (Speech Levels). Table 2 details the distribution of reasons.

*Table 1. Satisfaction Results*

<b>Item</b>	<b>Satisfaction Level Percentage</b>	
Item 1 - Syntax	I am satisfied.	0%
	I am dissatisfied.	100%
Item 2 - Speech Levels	I am satisfied.	0%
	I am dissatisfied.	100%

*Table 2. Reasons for Dissatisfaction*

<b>Item</b>	<b>Reason for Dissatisfaction</b>	<b>Percentage</b>	<b>Category</b>
Item 1 - Syntax	a. The content of the prompt distracted me.	32%	Clarity
	b. I do not understand what type of error I should mention.	27%	Clarity

Item	Reason for Dissatisfaction	Percentage	Category
Item 2 - Speech Levels	c. It does not correspond to what we studied in class.	7%	Relevance
	d. This prompt makes me feel uncomfortable.	34%	Relevance
	a. The content of the prompt distracted me.	10%	Clarity
	b. I do not understand what type of error I should mention.	15%	Clarity
	c. It does not correspond to what we studied in class.	3%	Relevance
	d. This prompt makes me feel uncomfortable.	72%	Relevance

### Item-level profiles

Item 1- Syntax: A combined 59% of dissatisfied responses cited clarity-coded reasons (content distracts, 32%; I don't understand what error I should mention, 27%), while 34% selected this prompt makes me feel uncomfortable and 7% pointed to outside our class scope. This pattern indicates two concurrent mechanisms: (a) opaque task cues about the precise operation expected (naming a specific syntactic fault), and (b) affective load introduced by a violent scenario. Both mechanisms are classic sources of construct-irrelevant variance that deflect attention from the intended metalinguistic task (Cassany, Luna, & Sanz, 2007; Fumero, 2021; Ogange et al., 2018).

Item 2- Speech levels: Here, 72% of dissatisfied students selected the security-coded reason (this prompt makes me feel uncomfortable), with smaller clarity-coded percentages (content distracts, 10%; uncertain about the error, 15%) and a minimal relevance signal (3%). The wording's explicit appeal to "vulgar speech" likely activated stigma cues that students interpreted as socially indexical rather than analytically neutral—a dynamic long noted in sociolinguistic accounts of attitude and identity (Trudgill & Hernández Campoy, 2007) and in applied-linguistics warnings about prescriptive labels that overshadow the learning target (Cassany et al., 2007).

### Interpreting the patterns

Across items, the clarity vs. security balance differs in degree but not in kind:

- Clarity dominated Item 1 (59%), signaling the need to name the requested operation (e.g., "identify the syntactic fault") and to stabilize the error label expected in the response.
- Security/affective load dominated Item 2 (72%), indicating that stigmatizing terminology can overwhelm the analytic objective and depress performance even when students know the content (Fumero, 2021; Ogange et al., 2018).

These results empirically support the theoretical stance in Sections 1–2: micro-features of wording – task verbs, scenario framing, and register labels– can either scaffold or undermine student performance in online tests. They also validate the utility of a brief, course-embedded perception check to surface where prompts compromise linguistic security and linguistic neutrality before grades are affected.

### **Improving actions derived from the evidence**

Guided by Tables 1-2 and by the literature (Cassany et al., 2007; Trudgill & Hernández Campoy, 2007; Fumero, 2021; Ogange et al., 2018), we implemented the following edits:

- Clarify the requested operation (Item 1). Make the task verb explicit (e.g., “Identify the syntactic fault in the sentence and name the category”). Provide a neutral exemplar of the expected label during instruction to reduce guesswork about nomenclature.
- Neutralize affective content (Item 1). Replace the violent scenario with a discipline-neutral sentence that still contains the same grammatical issue. This limits affective load while preserving cognitive demand.
- Reframe the register target (Item 2). Retire the stigmatizing label “vulgar speech” and recast the task as register matching or editing from informal to formal in a short excerpt. This keeps the analytic focus on fit to context rather than on social judgments.
- Calibrate the rubric (both items). Separate points for content mastery (recognizing the error/register) from points for decoding the prompt. When wording has demonstrably interfered, the rubric should prevent double penalization.

### **Portability and institutional use**

Because the instrument is brief and the analytics are descriptive, the approach is readily portable: any course relying on short prompts can embed a two-step perception check, examine clarity vs. security signals, and revise/improve wording before summative use. At the institutional level, results can be logged as part of routine assessment cycles, documenting how small changes in prompt design yield fairer evidence of learning without diluting rigor.

In sum, Tables 1-2 provide convergent evidence that wording –through clarity cues and affective load– was the principal source of difficulty in the two prompts. The edits above translate those signals into concrete, low-cost revisions aligned with applied-linguistics guidance on text design (Cassany et al., 2007) and sociolinguistic insights on stigma and reception (Trudgill & Hernández Campoy, 2007), while addressing the psycholinguistic risks of non-neutral content highlighted in assessment research (Fumero, 2021; Ogange et al., 2018).

## **Conclusion**

This proceedings contribution shows that a brief, course-embedded perception check can uncover wording-related barriers that compromise linguistic security in online basic-writing tests. The item-level profiles –clarity pressures in the syntax prompt and affective load in the speech-levels prompt– translate directly into small, testable edits: make the requested operation explicit, neutralize stigmatizing or emotionally charged content, and calibrate rubrics so that decoding the prompt is not conflated with mastery of linguistic content. Framed as revising/improving assessment design, this approach aligns with student-centered evidence practices and continuous-improvement cycles advocated in institutional assessment work (Jankowski et al., 2018; Kumar & Rani, 2021).

The modality matters. In prior face-to-face sections, the same targets met expected thresholds, whereas online delivery amplified students' dependence on literal wording and reduced opportunities for in-the-moment clarification. This aligns with distinctions between well-designed online learning and emergency or highly constrained remote conditions, where channel limitations heighten sensitivity to micro-features of text (Hodges et al., 2020; Meyer & Murrell, 2021). Future studies should therefore compare prompt variants across modalities, triangulate perception data with direct evidence of learning, and incorporate brief cognitive interviews to validate whether revised wording actually reduces construct-irrelevant variance.

For institutional use and portability, the protocol is intentionally lightweight: embed a two-step perception check after targeted items, code reasons into clarity/relevance/security, enact rapid wording edits, and communicate the changes to students as part of routine assessment reporting. Documenting these micro-revisions within program assessment plans strengthens transparency and supports equity goals while preserving cognitive demand (Jankowski et al., 2018; Kumar & Rani, 2021). In short, revising prompt language is a low-cost, high-leverage action that institutions can adopt broadly to enhance fairness and validity in online assessment (Hodges et al., 2020; Meyer & Murrell, 2021).

## **Acknowledgements**

All acknowledgments (if any) should be included in a separate section before the references and may include list of peoples who contributed to the work in the manuscript but not listed in the author list. If your research was supported by a Research Grant, you may include details here. This section is optional.

## **Declaration of Interest Statement**

The author declares that he has no conflict of interests.

## References

- Bawa, P. (2020). Retention in online courses: Exploring issues and solutions—A literature review. *SAGE Open*, 10(1), 1-12. <https://doi.org/10.1177/2158244019897893>
- Cassany, D.; Luna, M.; & Sanz, G. (2007). Enseñar lengua. Graó.
- Fumero, J. (2021). El impacto del inglés como vehículo de difusión científica en la universidad española. *Revista Española de Lingüística Aplicada*, 23, 111-128.
- García, D., & Fumero, M. (2010). *Tendencias en lingüística general y aplicada*. Frankfurt: Peter Lang.
- Trudgill, P., & Hernández Campoy, J. M. (2007). *Diccionario de sociolingüística*. Madrid: Editorial Gredos.
- Herrera Santana, J. (Coord.), & Díaz Galán, A. (Coord.). *Aportaciones al estudio de las lenguas: Perspectivas teóricas y aplicadas*. Romance Studies.
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Ibarra-Sáiz, M.S., Lukas-Mujika, J.-F., Ponce-González, N., & Rodríguez-Gómez, G. (2023). Percepción del profesorado universitario sobre la calidad de las tareas de evaluación de los resultados de aprendizaje. *RELIEVE: Revista Electrónica de Investigación y Evaluación Educativa*, 29(1). <https://doi.org/10.30827/relieve.v29i1.27404>
- Jankowski, N.A., Timmer, J.D., Kinzie, J., & Kuh, G.D. (2018). *Assessment that matters: Trending toward practices that document authentic student learning*. National Institute for Learning Outcomes Assessment.
- Kumar, V., & Rani, S. (2021). A comparative study of assessment methods in online and traditional classrooms: Implications for higher education. *International Journal of Educational Management*, 35(5), 1029-1044. <https://doi.org/10.1108/IJEM-01-2020-0012>
- Mace, M.K., & Pearl, D. (2020). Rubric development and validation for assessing comprehensive internationalization in higher education. *International Journal of Educational Management*, 25(1), Article 9986579. <https://doi.org/10.1177/1028315319865790>
- Meyer K.A., & Murrell D. (2021). Assessing student learning in online environments: A review of the literature on best practices and challenges. *Online Learning*, 25(2), 1-20.
- Mintz S. (2020). New approaches to assessing institutional effectiveness: How to ensure that institutions improve instructional quality and effectiveness and enhance equity and academic postgraduation outcomes. *Inside Higher Ed*. <https://www.insidehighered.com/blogs/highered-gamma/new-approaches-assessing-institutional-effectiveness>

- Ogange B.O., Agak J.O., Okelo K.O., & Kiprotich P. (2018). Student perceptions of the effectiveness of formative assessment in an online learning environment. *Open Praxis*, 10(1), 29–39.
- Seidman I. (2019). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (5th ed.). Teachers College Press.
- Suskie L. (2018). *Assessing student learning: A common-sense guide*. Wiley.
- Volkwein J.F. (2010). Assessing student outcomes: Why, who, what, how? New Directions for Institutional Research; Assessment Supplement 2009.
- Volkwein J.F. (2011). Gaining ground: The role of institutional research in assessing student outcomes demonstrating institutional effectiveness. New Directions for Institutional Research; Assessment Supplement 2010.