

*Heritage, second and third language learner  
processing of written corrective feedback:  
Evidence from think-alouds*

Melissa A. Bowles

University of Illinois at Urbana-Champaign, USA

<https://orcid.org/0000-0003-1896-8813>

[bowlesm@illinois.edu](mailto:bowlesm@illinois.edu)

Kacie Gastañaga

University of Illinois at Urbana-Champaign, USA

<https://orcid.org/0000-0001-8756-2687>

[kacielg2@illinois.edu](mailto:kacielg2@illinois.edu)

Abstract

This study compares the processing of three different types of written corrective feedback (WCF) by heritage language (HL), second language (L2), and third language (L3) learners who wrote and revised three short essays and received a different type of WCF for each essay (i.e., direct, coding, or underlining). Comparison of pre- and post-feedback texts and analysis of think-alouds served as the basis for determining whether one type of feedback promoted higher depth of processing (DoP) and whether this processing was mediated by error type and language background. The findings indicate that feedback type did interact with DoP, and that this interaction was in some ways mediated by learner background and error type. This research serves as a first step toward understanding how these three learner groups are impacted by these commonly used feedback types and is therefore important to drive evidence-based pedagogical decisions.

*Keywords:* written corrective feedback; depth of processing; think-alouds; heritage language learners; third language learners

## 1. Introduction

Written corrective feedback (WCF) is “any explicit attempt to draw a learner’s attention to a morphosyntactic or lexical error” (Polio, 2012, p. 375) in writing. The focus of WCF research has traditionally been on the accuracy of the final written product, but in recent years the processes learners engage in when they receive feedback have begun to receive empirical attention based on the recognition that without information about how WCF was processed, it is difficult to ascertain anything more than whether a revision was made (see Roca de Larios & Coyle, 2022, for a recent review).

Leow’s (2020) feedback processing framework is the most recent account of how engaging with WCF may (or may not) lead to changes in learners’ linguistic systems. Specifically, the framework explains that once learners have received feedback, they must minimally pay attention to it for it to potentially be converted into intake. Feedback processing pertains to how the learner cognitively processes the feedback in relation to current linguistic knowledge. If the feedback is further processed, whether with a low or high DoP, there is potential for previously learned knowledge to be restructured. The new restructured information (which might or might not be fully accurate at this point) can then replace or combine with the original knowledge in the learner’s internal system. Leow notes that it is possible for the learner to retain both accurate and inaccurate items in the system, and only measures of delayed performance can indicate whether a complete, accurate restructuring has taken place (resulting in system learning), or whether such restructuring was fleeting, as evidenced by accuracy immediately after feedback provision but regression to previous inaccurate knowledge at a later time.

Our research is couched in Leow’s (2020) framework and investigates the processing and impact of WCF on three different learner populations – second language (L2) learners, third language (L3) learners, and heritage language (HL) learners of Spanish in a university-level Spanish course in which composition is one component of the four-skills curriculum. These three populations are theoretically interesting, with heritage language learners understudied in WCF and L3 learners never having been the focus of prior WCF processing studies. Furthermore, examining the three populations is an ecologically valid choice since all of them are enrolled in the same classes. In what follows we first review different types of WCF and then move on to describing previous research findings on feedback processing and learner profiles as the necessary background to situate our study.

## 2. Literature review

### 2.1. Types of WCF

WCF differs along several dimensions, including its focus (i.e., limited to a restricted set of linguistic elements or comprehensive), its timing (i.e., synchronous or asynchronous with writing), and its explicitness (i.e., direct or indirect). For our study we chose asynchronous, comprehensive feedback because of ecological validity considerations in our context and because the potential effectiveness of comprehensive WCF has attracted less empirical attention than focused WCF (Cerezo et al., 2019, p. 174). To benefit from implicit (indirect) feedback, learners need to have enough proficiency in the language to be able to understand the nature of their errors, whereas explicit (direct) correction can be suitable for beginners as the feedback provides correction (Kang & Han, 2022). We chose to include both direct WCF and two types of indirect WCF – underlining and metalinguistic coding – as they are commonly used in language classrooms (Zhang et al., 2021) and for their theoretical suitability for learners of intermediate proficiency. In terms of processing, as Kang and Han (2022) explain, “explicit WCF places the processing responsibility in the hands of the feedback giver (i.e., the teacher), while implicit WCF shifts the responsibility to the feedback receiver (i.e., the learner)” (p. 221). We therefore anticipated that direct feedback might be processed less deeply than indirect feedback.

### 2.2. Findings of previous studies on the processing of WCF

Previous research has used two main introspective methods to examine how learners process feedback – written languaging (Cerezo et al., 2019; Manchón et al., 2020; Manchón & Roca de Larios, 2011; Suzuki, 2012, 2017) and think-alouds (Caras, 2019; DeRobles, 2019; Kim & Bowles, 2019; Park & Kim, 2019; Qi & Lapkin, 2001; Sachs & Polio, 2007; Suh, 2010). Leow and Manchón (2022) have convincingly argued for the affordances of think-alouds, stating that “think-aloud protocols, although a time-consuming undertaking mostly due to transcription, segmentation, and coding, arguably elicit the richest data on cognitive processes employed during task performance” (p. 306). Our study therefore adopted think-alouds for this purpose and we turn now to reviewing existing process research that has used this data collection procedure.

The earliest studies to examine WCF processing using think-alouds are Qi and Lapkin (2001) and Sachs and Polio (2007), both of which involved ESL learners who received reformulations. Although neither used the metric of DoP per se, in Qi and Lapkin (2001) “substantive” rather than “perfunctory” comments

were related to an improved written product and in Sachs and Polio (2007) errors accompanied by metalanguage or a hypothesis were more likely to be corrected in the revision. Park and Kim (2019), who examined processing of indirect feedback, similarly found that deeper processing resulted in better ability to self-correct marked errors. Where think-aloud studies comparing the processing of multiple types of WCF are concerned, direct feedback has been associated with lower levels of processing than reformulations and indirect feedback (Caras, 2019; Kim & Bowles, 2019). Suh (2010), however, found in her study that the intermediate-level ESL learners who received direct feedback manifested *higher* levels of awareness and showed significantly greater learning gains than those who received indirect feedback (metalinguistic coding) on the past counterfactual conditional.

It seems that the relationship between feedback and its processing may depend on a number of factors, including the L2 writer's proficiency level, the nature of the errors, and the extent to which the indirect feedback provides clues as to the nature of the error. For instance, Caras (2019) found that for preterit/imperfect errors, but not for *ser/estar* errors, most L2 Spanish participants who received metalinguistic feedback processed at a high level. Looking further into the relationship between error types and DoP, Kim and Bowles (2019) found that errors related to sentence structure and organization were associated with high DoP, whereas punctuation, word choice, and tense (surface level errors) were associated more often with low DoP. For Caras' beginning-level learners, the indirect feedback type that provided the fewest clues as to the nature of the error (crossing it out) generally led to low level processing, most likely because learners did not have enough target language knowledge to understand what was incorrect about their writing. In contrast, the indirect, metalinguistic coding feedback generally led to high level processing, presumably because learners relied on the code to identify the source of the error.

As this brief review shows, our understanding of the factors that impact WCF processing for L2 learners is just beginning to emerge and further research is needed to disentangle the roles learners play. All of the aforementioned studies examined how L2 learners of different languages, contexts, and proficiencies processed WCF. Other populations have featured less prominently in research. For instance, just one study (DeRobles, 2019) investigated how heritage learners process WCF. This study compared how high and "very high" proficiency learners processed direct and indirect (metalinguistic coding) feedback. Deeper processing was associated with greater accuracy between drafts, particularly for the metalinguistic coding group. Again, relationships were not straightforward, as direct feedback led to improved accuracy for some lower proficiency learners on certain error types, and proficiency seemed to mediate the relationship between WCF and DoP.

As studies of WCF processing are still scarce for HL learners and, to our knowledge, nonexistent for L3 learners, it is important to compare the feedback processing behavior of those learners who so often coexist in the foreign or second language classroom yet have diverse abilities and needs due to their unique linguistic backgrounds, which we discuss next. This expansion of the populations under study will likely contribute not only to expanding the empirical basis, but also to assessing whether available research findings can be generalized across populations and contexts.

### 2.3. HL, L2, and L3 learners

US college and university language classrooms are more diverse than ever (Association of American Colleges and Universities, 2019), with Spanish classes particularly so because it is the most widely taught language other than English in the country (Looney & Lusin, 2019). In many classrooms, L2 learners of Spanish who were raised monolingually in English coexist with heritage learners of Spanish, defined by Valdés (2000) as learners that grew up hearing and using the language at home. At the institution where this study was carried out, there is also a growing number of L3 learners of Spanish, who were raised speaking a heritage language such as Russian or Urdu at home, are dominant in English (the majority language), and choose to learn Spanish in a classroom setting for professional or personal reasons.

Previous research tells us that these learner types have different strengths and needs when it comes to language learning (e.g., Bowles & Montrul, 2017; Cenoz, 2013; Gatti & O'Neill, 2017; Gurzynski-Weiss, 2010). HL learners often have limited knowledge of grammar rules and metalinguistic terminology since they acquire the language naturalistically at home and rarely receive literacy instruction in Spanish at school in the U.S. However, they often have stronger pronunciation skills, greater oral fluency, and higher oral comprehension skills than L2 or L3 learners (Carreira & Kagan, 2011; Mrak, 2020). Past research has shown that HL learners tend to rely on their "ear" even when completing written tasks, often reading their writing and any feedback aloud to determine whether it "sounds right" (Yanguas & Lado, 2012; Zamora, 2022). Furthermore, their relative lack of familiarity with metalinguistic labels suggests they might be less able to benefit from indirect feedback (and especially coding feedback) than their peers who learned Spanish primarily in a classroom setting. L2 learners are mainly exposed to the target language in the classroom, and therefore usually have stronger explicit knowledge of grammar rules than HL learners, but weaker comprehension skills and oral fluency (Carreira & Kagan, 2011). They tend to be quite used to metalinguistic terms and are often very accurate at using them to

describe the Spanish language (Bowles, 2011). This tendency could make them more adept at using indirect feedback (especially metalinguistic error coding feedback) than their HL peers. L3 learners have the advantage of coming to the Spanish language classroom with two languages already under their belt, which, as past research has shown, contributes to higher overall metalinguistic awareness (Bialystok, 2001). This tendency, combined with their classroom exposure to Spanish, may predispose them to benefit the most of the three groups from indirect feedback.

Given their profiles, HL, L2 and L3 learners might be expected to process feedback differently and in empirically and pedagogically relevant ways. From the latter perspective, empirical findings could usefully inform our decisions about feedback provision, and in view of a limited number of studies with all but L2 learners, research is essential to drive evidence-based pedagogy for the three learner groups. In an attempt to add to previous research on feedback processing and gain new pedagogically relevant knowledge, the following research questions guided our study:

1. Is there a relationship between type of written corrective feedback (error coding, underlining, or direct correction) and DoP for HL, L2, or L3 learners?
2. Is there a relationship between DoP and error type for HL, L2, or L3 learners?
3. Is there a relationship between DoP and accurate error revision for HL, L2, or L3 learners?

### 3. Method

#### 3.1. Participants

The participants in this study were 35 students of Spanish from a large public university in the US Midwest. They ranged from low- to high-intermediate proficiency in Spanish as determined by an abridged version of the *Diploma de Español como Lengua Extranjera* (DELE), a Spanish proficiency test frequently used in linguistics research (e.g., Montrul 2004; Montrul & Slabakova 2003). The DELE contains 30 multiple-choice vocabulary questions and a 20-question cloze test of grammar and vocabulary. Test-takers can score 0-50 points, with scores from 0-29 regarded as low proficiency, 30-39 as intermediate and 40-50 as advanced. The learners also completed the *Bilingual Language Profile* (BLP) (Birdsong et al., 2012), which was used to gather information about their language backgrounds and to identify them as HL, L2, or L3 learners. According to self-reports from the BLP, on average participants in the HL group began learning Spanish at birth and English at three years old and reported using Spanish with their families 55% of the time, and with their friends 10% of the time. The L3 learners spoke

a wide variety of home languages including Mandarin Chinese, Hindi, Amharic, Japanese, Polish, Punjabi, and Ukrainian. Like the HL group, they reported learning English at three years old on average and using their home languages with family more than half of the time (68%) and with their friends about 30% of the time. L2 learners reported having been exposed only to English until learning Spanish in a classroom setting in adolescence, and they indicated they infrequently used Spanish outside of class.

The learners were drawn from a range of university Spanish courses, with HL learners reporting having studied Spanish for an average of four years, L3 learners for 4.3 years, and L2 learners for an average of 5.5 years. In total, there were 7 HL learners of Spanish, 9 L3 learners, and 19 L2 learners. Table 1 reports average DELE scores for each group. There were no statistically significant differences in proficiency between groups.

Table 1 DELE scores by group

Group	<i>N</i>	DELE Avg. ( <i>SD</i> )	Range
HL	7	26.86 (6.5)	16-33
L2	19	26 (7.4)	15-36
L3	9	28.67 (6.73)	19-37

### 3.2. Task

Each participant wrote and revised three 200-250-word essays in response to prompts about Spanish in the US (see Appendix). These prompts were chosen because they covered topics participants would have had personal experience with, were accessible to learners at this proficiency level, and related to topics covered in the curriculum.

Participants were given a week to write each essay in their own time. While writing, they recorded their screens using Zoom to ensure they were only using Microsoft Word spellcheck, [linguee.com](http://linguee.com) and [wordreference.com](http://wordreference.com). These websites were allowed as resources because they provide translations of individual words or phrases like “get ready” rather than full sentences with context and examples. Participants submitted their screen recordings with each essay; essays without screen recordings were not accepted.

After submitting every single essay, each participant met with the researcher or a research assistant approximately one week later via Zoom to revise their initial texts. During each session, the participant received their marked-up essay with one of three WCF types (error coding, direct feedback, underlining) and shared their screen while they edited and thought aloud. The next prompt was sent 1-2 days after each revision session and the process was repeated a

total of three times for each participant, with the order of prompts and WCF counterbalanced to ensure that any differences were not due to order effects or prompt attributes. For each revision session, one-third of the participants received each prompt and feedback type. Due to attrition, the final counts are not perfectly balanced. Table 2 below illustrates the number of participants for each prompt and feedback condition during each writing session.

Table 2 Counterbalancing of participants for each writing session

	Prompt	Number of participants	Error coding	Direct feedback	Underlining
Writing Session 1	Teaching of Spanish	10	4	3	3
	Resources	17	7	5	5
	Signage	8	3	1	4
Writing Session 2	Teaching of Spanish	15	3	5	7
	Resources	8	3	4	1
	Signage	12	4	4	4
Writing Session 3	Teaching of Spanish	10	1	4	5
	Resources	10	4	4	2
	Signage	15	6	5	4

### 3.3. Feedback types

The three feedback types that were chosen – direct corrective feedback, metalinguistic error coding, and underlining – were used because they are the main types of feedback that are given in the classroom and because they have been the focus of many previous studies (Zhang et al., 2021). Table 3 shows the same sentence marked with each feedback type.

Table 3 Feedback types

Direct	Muchas personas saben que <u>aprendiendo</u> aprender un lenguaje <u>cuándo</u> cuando ellos son jóvenes es más fácil que <u>aprendiendo</u> aprender una lengua en la escuela <u>secundario</u> secundaria. (Many people know that learning a language when they are young is easier than learning a language in high school.)
Coding	Muchas personas saben que aprendiendo(=) una lengua en la escuela secundario(AGR). (VT) Verb Tense, (WC)= Word Choice, (SP)= Spelling, (=) = Same as previous error, (AGR)= Agreement
Underlining	Muchas personas saben que <u>aprendiendo</u> un lenguaje <u>cuándo</u> ellos son jóvenes es más fácil que <u>aprendiendo</u> una lengua en la escuela <u>secundario</u> .

Following Caras' (2019) claim that comprehensive WCF is ecologically valid since it reflects the way instructors tend to revise compositions in the classroom (p. 186), we elected to provide this type of feedback. The errors that were marked were article errors (including missing articles and issues with definite/indefinite articles), inflectional morphology, such as gender and subject verb agreement, mood, punctuation, prepositions, spelling errors, and sentence structure errors such as incomplete sentences, verb tense, and word choice. These error types were chosen because they encompass the majority of errors that language learners at this level make when writing.



To ensure the essays were being coded reliably, both authors provided feedback on 10% of the essays, divided equally among the three types of WCF. Interrater agreement on these essays was 94%, which was considered high enough that one author then coded the remaining essays.

### 3.4. Think-aloud protocols

The think-aloud protocol was designed following Bowles (2010), beginning with instructions and think-aloud practice for participants before they moved on to the revision task. The instructions asked learners to think aloud in English, Spanish, or a combination of the two as was most natural while they revised, saying what went through their mind as they reviewed their feedback. Next, the participants completed a brief think-aloud practice activity made up of three sentences with grammatical errors in English that had been marked with the type of feedback that corresponded to that day's session. Examples of each are included in Table 4.

Table 4 Think-aloud practice

---

Example sentence (Underlining feedback)	<u>They're wasn't</u> any students I knew at the school event.
Example think-aloud comments	Ok, so I have an issue with "they're." Maybe I spelled it wrong since there are many ways to spell it. And a problem with wasn't, too. Maybe it is supposed to be "weren't" since there is more than one student?
Practice sentence	He didn't <u>went</u> to the game yesterday.

---

After the think-aloud practice session, participants received their marked-up essay, shared their screen, and began to revise. The researcher started recording the Zoom session, turned off her video and muted her microphone, only speaking to remind the participants to continue to think aloud if they fell silent for more than a few seconds. This prompting was rarely necessary. Participants spent on average 15-20 minutes revising each essay.

### 3.5. Coding for DoP

The think-aloud recordings were transcribed verbatim. Participants' comments for each error were classified as high, medium, or low DoP, adopting the coding scheme used by Caras (2019), an adaptation of Leow's (2015) coding system. Table 5 provides descriptors for each category in the coding system and Table 6 contains examples of each type of DoP in each feedback condition.

Table 5 DoP coding system

Description	Low DoP	Medium DoP	High DoP
Descriptors	Shows no potential for processing target	Comments on target in relation to meaning	Arrives at an inaccurate, partially, or fully accurate underlying rule
	Repeats target or their production in the composition	Spends a bit more time processing target	Makes hypotheses regarding target
	Translates the phrase into English	Makes comments that indicate some processing of target	Provides an inaccurate, accurate and/or partially accurate rule
	Says what they mean in English	Refers to a previous instance with the same target	Corrects previous translation
	Repeats target or their production in the composition	Some level of cognitive effort to process target	Spends much time processing target
	Says they aren't sure what is wrong		High level of cognitive effort to process target
	Comments on the target but does not indicate any further processing		
	Avoids the correction		
	Does not spend much time processing target		
	Low level of cognitive effort to process target		

Table 6 DoP coding examples

High DoP	Umm . . . pero . . . "VT" . . . , verb tense. Okay . . . oh, that's because . . . I believe it's . . . paré? Conjugations (Looks up on Wordreference.com) . . . umm paré, sí ( <i>Coding</i> ). And then. Americana. Yeah, porque I was talking in the the feminine primera generación cause generación is feminine, so I have to change this to an "a." Ooh ( <i>Direct</i> ). Aquellos que son bilingües están mejor enfocan. (Pause). Están . . . Then this one probably too. Okay, they practice. Wait and so, those who are bilingual are better at concentrating because they practice. That should be practican. Pra-ti-can. Conjugate that. Because they practice. I think that's right, okay. Mejor . . . están . . . aquellos que son bilingües están mejor enfocan ( <i>Underlining</i> ). No es tan bueno, "AGR," agreement. Buena. ("It's not as good (feminine). AGR. Agreement. Good (masculine)") ( <i>Coding</i> ).
Medium DoP	Lenguaje again . . . guess you can't use that word ( <i>Direct</i> ). Muchas investigaciones creen que los hispanos . . . seria. Oh. Serian. ("Many studies believe that Hispanics . . . would be (singular). Oh. Would be (plural)") ( <i>Underlining</i> ). Tenido más práctica hablando ("Had more practice speaking") ( <i>Coding</i> ).
Low DoP	A los . . . ugh ohh ("To the . . . Ugh ohh") ( <i>Direct</i> ). Encan-encantar ("Lo-love") ( <i>Underlining</i> ).

Each researcher coded 10% of the transcripts separately and interrater agreement for DoP was 94%. The two authors discussed the disagreements and further elaborated the coding scheme accordingly, then one author coded the rest of the data. Each think-aloud comment was coded for high, medium, or low DoP, error type, and whether the revision was correct, incorrect (but changed from the original), or unchanged. In total, there were 1,910 comments across all three feedback conditions. Each participant provided an average of 54 think-aloud comments ( $SD = 27$ ). Descriptive statistics were used to answer the research questions in the following section.

#### 4. Results

The first research question asked about the relationship between the type of written corrective feedback and DoP for HL, L2, and L3 learners. Tables 7-9 illustrate each group's percentage of high, medium, and low DoP in each feedback condition. For all three learner groups, direct feedback was most often processed with low DoP (in 60% of cases). Error coding and underlining feedback, on the other hand, were processed with medium or high DoP approximately 60% of the time by HL and L2 learners. L3 learners also processed underlining feedback with medium or high DoP at this rate, but not error coding. These findings echo previous work which has found that indirect feedback leads L2 and HL learners to process more deeply than its direct counterpart (Caras, 2019; DeRobles, 2019; Kim & Bowles, 2019) and provide the first measures of L3 feedback processing. The two indirect feedback types were processed with similar rates of low (over 40%), medium (around 30%), and high DoP (around 30%) by HL and L2 learners. The L3 learners differed from the other groups in their processing of coding feedback. They had the lowest rates of high DoP for coding (17%) of the three groups, as compared to HL learners (26%) and L2 learners (32%). In addition, the L3 learners processed the most implicit type of feedback, underlining, with medium or high DoP 63% of the time, 7% more often than the L2 group and 15% more often than HL learners.

Table 7 HL DoP by feedback type

DoP	Code	Direct	Underline	Totals
Low	53 (45%)	92 (66%)	45 (42%)	190 (52%)
Med	35 (29%)	13 (9%)	32 (30%)	80 (22%)
High	31 (26%)	35 (25%)	30 (28%)	96 (26%)
Totals	119	140	107	366

Table 8 L2 DoP by feedback type

DoP	Code	Direct	Underline	Totals
Low	146 (42%)	203 (62%)	13 (44%)	480 (49%)
Med	89 (26%)	53 (16%)	85 (29%)	227 (23%)
High	112 (32%)	74 (22%)	80 (27%)	266 (27%)
Totals	347	330	296	973

Table 9 L3 DoP by feedback type

DoP	Code	Direct	Underline	Totals
Low	92 (48%)	108 (57%)	69 (37%)	269 (47%)
Med	66 (35%)	32 (17%)	53 (28%)	151 (27%)
High	32 (17%)	49 (26%)	67 (35%)	148 (26%)
Totals	190	189	189	568

Table 10 HL DoP by error type

Feedback type	DoP	ART	IM	M	P	PR	S	SS	T	WC
Code	Low	6 (46%)	9 (45%)	1 (33%)	4 (50%)	5 (56%)	15 (38%)	1 (50%)	4 (80%)	8 (42%)
	Med	4 (31%)	7 (35%)	1 (33%)	3 (37%)	2 (22%)	13 (33%)	0 (0%)	0 (0%)	5 (26%)
	High	3 (23%)	4 (20%)	1 (33%)	1 (12%)	2 (22%)	11 (28%)	1 (50%)	1 (20%)	6 (32%)
Direct	Low	9 (69%)	5 (63%)	2 (50%)	1 (20%)	13 (93%)	35 (70%)	4 (80%)	6 (60%)	17 (55%)
	Med	2 (15%)	0 (0%)	2 (50%)	2 (40%)	1 (7%)	2 (4%)	0 (0%)	1 (10%)	3 (10%)
	High	2 (15%)	3 (37%)	0 (0%)	2 (40%)	0 (0%)	13 (26%)	1 (20%)	3 (30%)	11 (35%)
Underline	Low	2 (22%)	5 (38%)	2 (33%)	4 (57%)	5 (42%)	12 (37%)	1 (50%)	5 (71%)	9 (47%)
	Med	4 (44%)	3 (23%)	4 (66%)	2 (29%)	4 (33%)	8 (25%)	0 (0%)	1 (14%)	6 (32%)
	High	3 (33%)	5 (38%)	0 (0%)	1 (14%)	3 (25%)	12 (37%)	1 (50%)	1 (14%)	4 (21%)

Note. ART = article, IM = inflectional morphology, M = mood, P = punctuation, PR = preposition, S = spelling, SS = sentence structure, T = tense, WC = word choice

Table 11 L2 DoP by error type

Feedback type	DoP	ART	IM	M	P	PR	S	SS	T	WC
Code	Low	17 (50%)	17 (32%)	5 (38%)	6 (60%)	19 (59%)	21 (51%)	9 (53%)	13 (46%)	39 (33%)
	Med	8 (24%)	19 (36%)	3 (23%)	2 (20%)	7 (22%)	8 (20%)	4 (23%)	4 (14%)	34 (28%)
	High	9 (26%)	17 (32%)	5 (38%)	2 (20%)	6 (19%)	12 (29%)	4 (23%)	11 (39%)	46 (39%)
Direct	Low	22 (50%)	35 (59%)	8 (57%)	7 (87%)	21 (55%)	28 (76%)	13 (68%)	14 (54%)	55 (65%)
	Med	7 (16%)	6 (10%)	5 (38%)	0 (0%)	10 (26%)	5 (13%)	3 (16%)	3 (11%)	14 (16%)
	High	15 (34%)	18 (31%)	1 (7%)	1 (12%)	7 (18%)	4 (11%)	3 (16%)	9 (35%)	16 (19%)
Underline	Low	16 (50%)	13 (32%)	2 (25%)	8 (73%)	11 (52%)	23 (38%)	13 (54%)	10 (38%)	35 (47%)
	Med	13 (41%)	12 (30%)	2 (25%)	2 (18%)	5 (24%)	19 (32%)	6 (25%)	9 (35%)	17 (23%)
	High	3 (9%)	15 (37%)	4 (50%)	1 (9%)	5 (24%)	18 (30%)	5 (21%)	7 (27%)	22 (30%)

Note. ART = article, IM = inflectional morphology, M = mood, P = punctuation, PR = preposition, S = spelling, SS = sentence structure, T = tense, WC = word choice

Table 12 L3 DoP by error type

Feedback type	DoP	ART	IM	M	P	PR	S	SS	T	WC
Code	Low	13 (59%)	19 (50%)	2 (25%)	1 (100%)	5 (38%)	20 (55%)	2 (100%)	2 (13%)	28 (52%)
	Med	3 (14%)	18 (47%)	5 (62%)	0 (0%)	8 (61%)	8 (22%)	0 (0%)	8 (53%)	16 (30%)
	High	6 (27%)	1 (3%)	1 (12%)	0 (0%)	0 (0%)	8 (22%)	0 (0%)	5 (33%)	10 (18%)
Direct	Low	22 (76%)	13 (42%)	3 (43%)	3 (100%)	14 (82%)	19 (63%)	4 (50%)	4 (33%)	26 (50%)
	Med	2 (7%)	6 (19%)	4 (57%)	0 (0%)	2 (12%)	4 (13%)	1 (12%)	4 (33%)	9 (17%)
	High	5 (17%)	12 (39%)	0 (0%)	0 (0%)	1 (6%)	7 (23%)	3 (38%)	4 (33%)	17 (33%)
Underline	Low	14 (67%)	12 (36%)	3 (25%)	0 (0%)	8 (40%)	20 (45%)	1 (14%)	3 (19%)	8 (22%)
	Med	4 (19%)	7 (21%)	5 (42%)	0 (0%)	6 (30%)	13 (29%)	2 (29%)	4 (25%)	12 (33%)
	High	3 (14%)	14 (42%)	4 (33%)	0 (0%)	6 (30%)	11 (25%)	4 (57%)	9 (56%)	16 (44%)

Note. ART = article, IM = inflectional morphology, M = mood, P = punctuation, PR = preposition, S = spelling, SS = sentence structure, T = tense, WC = word choice

The second research question asked if there was a relationship between DoP and error type for the three groups. Tables 10-12 include detailed data for each learner group. Direct corrective feedback led to the lowest DoP for most error types for all groups. That said, the HL group processed the few punctuation errors they made most deeply in the direct feedback condition, often spending considerable time reading and re-reading the sentence aloud to see if the corrected punctuation sounded right. The L2 learners processed article and preposition

errors with equally low DoP in every condition, most often editing them by simply saying the correction aloud (e.g., “*e/ país*” instead of just “*país,*” or “*por*” instead of “*para*”). Underlining feedback promoted the highest DoP for the majority of error types for all groups, followed closely by coding feedback for L2 learners. Specifically, L3 learners processed underlining feedback with the highest DoP for eight of the nine error types, the only exception being article errors, which were processed most deeply with coding feedback. L2 and HL learners demonstrated the most medium and high DoP for inflectional morphology, mood, preposition, and spelling errors when they received underlining feedback. The L2 learners seemed to benefit from the error coding feedback more than the other groups for punctuation, sentence structure, tense, and word choice errors as this feedback type promoted the highest DoP for these error types, most likely because it indicates the source of the error, something that might not be clear from underlining for these errors, especially to L2 learners whose knowledge is more rule-based than intuitive and who do not have heightened metalinguistic awareness like the L3 learners.

Table 13 HL DoP and error revision

DoP	Accurately corrected			Unchanged			Not accurately corrected		
	Code	Direct	Under-line	Code	Direct	Under-line	Code	Direct	Under-line
Low	40 (42%)	82 (63%)	21 (33%)	4 (80%)	0 (0%)	6 (67%)	9 (50%)	10 (91%)	18 (53%)
Med	29 (30%)	13 (10%)	18 (28%)	1 (20%)	0 (0%)	2 (22%)	5 (28%)	0 (0%)	12 (35%)
High	26 (27%)	34 (26%)	25 (39%)	0 (0%)	0 (0%)	1 (11%)	4 (22%)	1 (9%)	4 (12%)
Totals	95	129	64	5	0	9	18	11	34
Overall Totals	288/365 (79%)			14/365 (4%)			63/365 (17%)		

Table 14 L2 DoP and error revision

DoP	Accurately corrected			Unchanged			Not accurately corrected		
	Code	Direct	Under-line	Code	Direct	Under-line	Code	Direct	Under-line
Low	85 (35%)	198 (62%)	68 (38%)	20 (77%)	2 (67%)	21 (60%)	41 (52%)	4 (57%)	42 (50%)
Med	67 (27%)	50 (15%)	49 (27%)	2 (8%)	1 (33%)	11 (31%)	20 (26%)	2 (26%)	25 (30%)
High	91 (37%)	73 (23%)	61 (34%)	4 (15%)	0 (0%)	3 (8%)	17 (22%)	1 (22%)	16 (19%)
Totals	243	321	178	26	3	35	78	7	83
Overall Totals	742/974 (76%)			64/974 (7%)			168/974 (17%)		

Table 15 L3 DoP and error revision

DoP	Accurately corrected			Unchanged			Not accurately corrected		
	Code	Direct	Under-line	Code	Direct	Under-line	Code	Direct	Under-line
Low	67 (47%)	105 (62%)	46 (34%)	9 (75%)	3 (50%)	8 (73%)	16 (43%)	0 (0%)	15 (34%)
Med	46 (33%)	20 (12%)	37 (28%)	2 (17%)	2 (33%)	2 (18%)	18 (49%)	0 (0%)	14 (32%)
High	28 (20%)	45 (26%)	51 (38%)	1 (8%)	1 (17%)	1 (9%)	3 (8%)	3 (100%)	15 (34%)
Totals	141	170	134	12	6	11	37	3	44
Overall Totals	445/588 (80%)			29/588 (5%)			84/588 (15%)		

The third research question concerned the relationship between DoP and error revision for each group. Tables 13-15 include the percentages of accurately

corrected, unchanged, and inaccurately revised errors in each group by feedback condition. Learners from all three groups were able to accurately revise approximately 80% of their errors, regardless of DoP or feedback type and rarely left errors unchanged. Unsurprisingly, the direct feedback condition allowed learners in each group to revise many errors correctly with low DoP. When errors were left unchanged, they were most often processed with low DoP, either because the learner simply skipped the error or took a quick look and decided they did not know how to correct it. Few errors were left unchanged by any group in the direct feedback condition since it provided learners with the correct revision. In each group roughly 15% of errors were incorrectly revised. This mostly occurred when the learners demonstrated low or medium DoP and happened most often in the underlining and coding feedback conditions. Notably, there were some instances in each group of high DoP leading to inaccurate revisions, meaning that higher DoP did not always lead to accurate revision and conversely, medium or high DoP was not always required for accurate revision. Correct revision accompanied by high DoP occurred most often in the underlining feedback condition for HL and L3 learners and in the error coding condition for L2 learners. Based on Leow (2020) and past research on the relationship between high DoP and subsequent learning outcomes (Leow, 2015), these are instances that have the greatest potential for leading to learning gains.

While high DoP does not always lead to correct revision and correct revision does not necessarily require high DoP, in general deeper processing has been found to be associated with more correct revision (Park & Kim, 2019). Since direct feedback provides learners with the correct answer, when we focus only on the error coding and underlining conditions, we see that higher DoP does appear to be linked more with accurate revisions and lower DoP is associated with unchanged or incorrectly revised errors. This pattern was clear for all three groups, where between 57-67% of the correctly revised errors were processed with medium or high DoP, 60-80% of the unchanged errors were processed with low DoP, and approximately 50% of the inaccurately revised errors in each group were processed with low DoP. Only the L3 learners in the underlining feedback condition broke this pattern as they were equally likely to process their incorrectly revised errors with low, medium, or high DoP.

## 5. Discussion

In this study we set out to explore the processing of three different types of written corrective feedback by HL, L2, and L3 learners with the ultimate aim of contributing new data on feedback processing from diverse learners of Spanish that could be useful empirically (in terms of the generalizability of findings on feedback

processing across populations and contexts) and pedagogically (especially in terms of pedagogical decision-making related to the provision of feedback).

Our first research question asked if there was a relationship between type of WCF and DoP for learners. Our results confirm that feedback type and depth of processing do interact. Across all groups, direct corrective feedback, the most transparent feedback type, led to the lowest DoP. Error coding and underlining feedback promoted medium or high DoP more than half the time for every learner group. These findings echo previous research which has found that indirect feedback tends to promote higher DoP so long as the learners have sufficient proficiency to understand the feedback (Caras, 2019; Kim & Bowles, 2019). The HL and L2 learners in this study had similar rates of high, medium, and low DoP for each feedback type, while the L3 data present very distinct patterns. That is, the L3 learners processed the error coding feedback much more shallowly than the other groups, only ever processing it with high DoP 17% of the time as opposed to 26% for HL learners and 32% for the L2 group. Conversely, they processed the underlining feedback more deeply than any other group. From our observations, the L3 learners were often able to resolve their errors very quickly upon seeing the error code, often simply stating the correct answer. Researchers have indicated that L3 learners have advanced metalinguistic skills due to their exposure to multiple languages (Bialystok, 2001), and it seems that the learners in this study may have been able to use such skills to their advantage with this feedback type. That said, due to the small sample size in this study, more data should be collected to determine the generalizability of this finding. Similarly, a larger sample size would be needed to confirm the general trend in our data that feedback type and depth of processing interact, and that this interaction is mediated by learner-related characteristics and background.

Our second research question concerned the relationship between DoP and error type. Overall, we found that error type interacted with DoP for these learner groups. Underlining feedback promoted the highest DoP for many error types including inflectional morphology, mood, prepositions, and spelling for all three learner groups, again echoing findings that indirect feedback promotes higher DoP and providing evidence of the way this can be mediated by error type for these learner groups. The L2 group demonstrated the highest DoP when they received error coding feedback for punctuation, sentence structure, tense, and word choice errors. Since L2 learners' linguistic knowledge is often more explicit and rule-based than that of HL learners, which is more implicit and intuitive, they may have benefited more from the direct metalinguistic guidance provided to them via the error coding system. The HL learners had the highest DoP for punctuation, tense, and word choice errors in the direct feedback condition. HL learners are known to use a "sound it out" approach when revising

written work (Yanguas & Lado, 2012; Zamora, 2022), and that is what we saw happening in this case. Many HL learners dedicated significant amounts of time to reading and rereading what they had written, making sure the added punctuation or suggested new word sounded right, whereas other groups were more likely to simply add a punctuation mark or change the word and move on. What this would mean in terms of learning outcomes is unclear and should be studied in future research. These data provide evidence of interaction between error type and DoP, yet it is still difficult to draw clear conclusions from these results due to small sample sizes for some error types. Overall, it seems that feedback type and learner background mediated DoP to a greater extent than error type.

Our third and final research question asked about the relationship between DoP and error revision. We found that DoP and error revision did interact, with higher DoP leading to more accurate error revision for all groups in the underlining and error coding feedback conditions, which aligns with previous findings about more substantive processing and error revision (Park & Kim, 2019; Qi & Lapkin, 2001; Sachs & Polio, 2007). Underlining feedback promoted the highest DoP and correct revision for the HL and L3 groups, while error coding best achieved this for L2 learners. The L3 group was able to revise correctly with low DoP more often than the HL or L2 groups. Again, this often occurred because they were able to quickly identify their mistakes and fix them without deliberating. That said, accurate revision and high DoP did not always go hand in hand. There were examples in every group and under every feedback condition of low DoP leading to accurate revisions and medium or high DoP leading to inaccurate revisions. These findings further reinforce the idea that writers' background and education experience are key variables to be considered when investigating feedback processing and its potential effects on the revised text produced. These insights also point to the necessary caution in generalizing findings across educational contexts and populations.

## 6. Conclusions and pedagogical implications

This exploratory study provides a first look at L3 processing of WCF and bolsters the amount of data we have on HL learners, a group that has been understudied to date. The analyses showed that DoP was mediated by feedback type for HL, L2, and L3 learners. Direct corrective feedback promoted the lowest average DoP for all three groups, while error coding and underlining promoted medium and high DoP more often. The HL and L2 learners demonstrated similar DoP patterns under each feedback condition, with coding and underlining promoting similar amounts of low, medium, and high DoP for each group, whereas the L3 group was unique in processing error coding with much lower DoP than the



other two groups and underlining with high DoP more often. It seems that these learners, potentially due to their increased metalinguistic awareness, found it easier to understand their errors marked with error coding than the other groups and also had an advantage when it came to interpreting and applying the implicit underlining feedback.

Along with feedback type, error type played a role in DoP for each learner group. Overall, underlining led to the highest DoP for most error types for L3 learners. Inflectional morphology, mood, prepositions, and spelling were processed most deeply in the underlining condition by all learners, but the L2 learners processed punctuation, sentence structure, tense, and word choice errors most deeply in the error coding condition and the HL group processed punctuation, tense, and word choice errors most deeply upon receiving direct corrective feedback, often sounding out the suggestion to see if they agreed with it before editing. Finally, higher DoP was associated with correct revision and lower DoP tended to be related to leaving errors unchanged or revising them incorrectly, at least with coding and underlining feedback types. Direct corrective feedback led all three learner groups to correctly revise their errors with low DoP far more often than the other two types of WCF due to its transparent nature.

While this investigation was primarily exploratory in nature due to the small sample size, a few pedagogical implications can be drawn. When it comes to feedback processing, underlining feedback promoted the highest DoP for the majority of error types in all learner groups. The L3 learners especially processed this feedback type more deeply than the other learner groups, indicating that this feedback type is likely ideal for them if the goal is to promote deep processing. The L2 group benefited more than the other two groups from the metalinguistic clues provided via the error coding feedback, most likely because their linguistic knowledge is primarily explicit and rule-based. That said, the HL and L2 groups had similar rates of DoP in the error coding and underlining conditions, so both feedback types could be beneficial for them. Direct corrective feedback promoted the lowest DoP for all groups across most error types. However, HL learners often spent significant amounts of time sounding out the direct feedback to make sure they agreed with it, something that the other two groups did not do.

Some limitations that should be considered are the small number of participants in the HL and L3 groups and the lack of a measure of learning. Our data shed light on how learners process WCF and what they do when they revise under different conditions, but we cannot speak to what impact the WCF had on their learning. Leow (2020) explains that higher DoP can potentially lead learners to restructure their inaccurate linguistic knowledge, and it is important to include long-term measures of learning to be able to make conclusions about the impact of WCF on language acquisition.

Future studies of feedback processing should continue to compare these three learner groups. Descriptions of observed learner revision behavior would also be beneficial. In this study we found that HL and L3 learners' behavior while revising differed from that of their L2 peers in some ways. L3 learners often revised very quickly and correctly, making minimal comments about their errors yet seemingly understanding the feedback target. Students in the HL group often took time to read and reread the feedback aloud, sounding out their mistakes and the suggested solutions, even with direct feedback. As most DoP coding schemes used to date were designed with L2 learners in mind and do not account for these behaviors, additional data should be collected to determine whether adaptations to DoP coding schemes may be in order.

## References

- Association of American Colleges & Universities. (2019, March). *Facts & figures: College students are more diverse than ever. Faculty and administrators are not.* <https://www.aacu.org/aacu-news/newsletter/2019/march/facts-figures>
- Bialystok, E. (2001). Metalinguistic aspects of bilingual processing. *Annual Review of Applied Linguistics*, 21, 169-181. <https://doi.org/10.1017/S0267190501000101>
- Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. COERLL, University of Texas at Austin. <https://sites.la.utexas.edu/bilingual/>.
- Bowles, M. A. (2010). *The think-aloud controversy in second language research.* Routledge. <https://doi.org/10.4324/9780203856338>
- Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, 33, 247-271.
- Bowles, M. A., & Montrul, S. (2017). Instructed heritage language acquisition. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 488-502). Routledge. <https://doi.org/10.4324/9781315676968-27>
- Caras, A. (2019). Written corrective feedback in compositions and the role of depth of processing. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 186-198). Routledge. <https://doi.org/10.4324/9781315165080-13>
- Carreira, M., & Kagan, O. (2011). The results of the national heritage language survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals*, 44(1), 44-64. <https://doi.org/10.1111/j.1944-9720.2010.01118.x>
- Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, 46(1), 71-86. <https://doi.org/10.1017/S0261444811000218>
- Cerezo, L., Manchón, R. M., & Nicolás-Conesa, F. (2019). What do learners notice while processing written corrective feedback? A look at depth of processing via written languaging. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 173-187). Routledge. <https://doi.org/10.4324/9781315165080-12>
- DeRobles, G. (2019). *The effects of type of written corrective feedback and level of proficiency on processing and accuracy in heritage language learners of Spanish.* (Doctoral dissertation, Georgetown University).
- Gatti, A., & O'Neill, T. (2017). Who are heritage writers? Language experiences and writing proficiency. *Foreign Language Annals*, 50(4), 734-753. <https://doi.org/10.1111/flan.12291>

- Gurzynski-Weiss, L. (2010). *Factors influencing oral corrective feedback provision in the Spanish foreign language classroom: Investigating instructor native/nonnative speaker status, SLA education, & teaching experience*. (Doctoral dissertation, Georgetown University). <https://repository.library.georgetown.edu/bitstream/handle/10822/553240/gurzynski-WeissLaura.pdf?sequence=1>
- Kang, E. Y., & Han, Z. (2022). Written corrective feedback: Short-term and long-term effects on language learning. In R. M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 213-225). Routledge. <https://doi.org/10.4324/9780429199691-23>
- Kim, H. R., & Bowles, M. (2019). How deeply do second language learners process written corrective feedback? Insights gained from think-alouds. *TESOL Quarterly*, 53(4), 913-938. <https://doi.org/10.1002/tesq.522>
- Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge. <https://doi.org/10.4324/9781315887074>
- Leow, R. P. (2020). L2 writing-to-learn: Theory, research, and a curricular approach. In R. M. Manchón (Ed.), *Writing and language learning: Advancing research agendas* (pp. 95-118). John Benjamins. <https://doi.org/10.1075/lllt.56.05leo>
- Leow, R. P., & Manchón, R. M. (2022). Directions for future research agendas on L2 writing and feedback as language learning from an ISLA perspective. In R. M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 299-311). Routledge. <https://doi.org/10.4324/9780429199691-31>
- Looney, D., & Lusin, N. (2019). *Enrollments in languages other than English in United States institutions of higher education, summer 2016 and fall 2016: Final Report*. Modern Language Association. <https://files.eric.ed.gov/fulltext/ED599007.pdf>
- Manchón, R. M., & Roca de Larios, J. (2011). Writing to learn in FL contexts: Exploring learners' perceptions of the language learning potential of L2 writing. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 181-207). John Benjamins. <https://doi.org/10.1075/lllt.31.13man>
- Manchón, R. M., Nicolás-Conesa, F., Cerezo, L., & Criado, R. (2020). L2 writers' processing of written corrective feedback: Depth of processing via written languaging. In W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching* (pp. 241-265). John Benjamins. <https://doi.org/10.1075/lllt.55.11man>
- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition*, 7, 125-142. <https://doi.org/10.1017/S1366728904001464>
- Montrul, S., & Slabakova, R. (2003). Competence similarities between native and near-native speakers: An investigation of the preterite-imperfect contrast in Spanish. *Studies in Second Language Acquisition*, 25, 351-398. <https://doi.org/10.1017/S0272263103000159>

- Mrak, A. (2020). Developing writing in Spanish heritage language learners: An integrated process approach. *Dimension*. <https://eric.ed.gov/?id=EJ1249861>
- Park, E. S., & Kim, O. Y. (2019). Learners' use of indirect written corrective feedback: Depth of processing and self-correction. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 214-228). Routledge. <https://doi.org/10.4324/9781315165080-15>
- Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing, 21*, 375-389. <https://doi.org/10.1016/j.jslw.2012.09.004>
- Qi, D., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing, 10*, 277-303. [https://doi.org/10.1016/S1060-3743\(01\)00046-7](https://doi.org/10.1016/S1060-3743(01)00046-7)
- Roca de Larios, J., & Coyle, Y. (2022). Learners' engagement with written corrective feedback in individual and collaborative L2 writing conditions. In R. M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 81-93). Routledge. <https://doi.org/10.4324/9780429199691-11>
- Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language Acquisition, 29*, 67-100. <https://doi.org/10.1017/S0272263107070039>
- Suh, B. R. (2010). *Written feedback in second language acquisition: Exploring the roles of type of feedback, linguistic targets, awareness and concurrent verbalization* (Doctoral dissertation, Georgetown University).
- Suzuki, W. (2012). Written languaging, direct correction and second language writing revision. *Language Learning, 62*, 1110-1133. <https://doi.org/10.1111/j.1467-9922.2012.00720.x>
- Suzuki, W. (2017). The effect of quality of written languaging on second language learning. *Writing & Pedagogy, 8*, 461-482. <https://doi.org/10.1558/wap.27291>
- Valdés, G. (2000). Introduction. In *Spanish for native speakers. AATSP professional development series handbook for teachers K-16, Volume 1* (pp. 1-20). Harcourt College.
- Yanguas, I., & Lado, B. (2012). Is thinking aloud reactive when writing in the heritage language? *Foreign Language Annals, 45*(3), 380-399. <https://doi.org/10.1111/j.1944-9720.2012.01198.x>
- Zamora, C. (2022). The secret is in the processing: Categorizing how heritage learners of Spanish process. In M. A. Bowles (Ed.), *Outcomes of university Spanish heritage language instruction in the United States* (pp. 81-102). Georgetown University Press.
- Zhang, T., Chen, X., Hu, J., & Ketwan, P. (2021). EFL students' preferences for written corrective feedback: Do error types, language proficiency, and foreign language enjoyment matter? *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2021.660564>

APPENDIX

Prompts

*Los recursos para los estudiantes*

- ¿Crees que las universidades en los Estados Unidos deberían tener recursos y programas específicamente para el apoyo de los estudiantes hispanos? ¿Por qué o por qué no? ¿Qué tipo de recursos serían ideales?  
(Do you think universities in the U.S. should have resources and programs specifically to support Hispanic students? Why or why not? What types of resources would be ideal?)

*La enseñanza de español*

- ¿Crees que en las escuelas públicas en los Estados Unidos se debería enseñar el español a todos los estudiantes desde el kínder, así como se enseña el inglés en otros países? ¿Por qué? ¿Te gustaría haber tenido esa oportunidad?  
(Do you think in public schools in the U.S. Spanish should be taught to all students starting in kindergarten the way English is taught in other countries? Why? Would you have liked to have that opportunity?)

*Las señales*

- ¿Crees que todas las señales de tránsito y los letreros en lugares públicos en los Estados Unidos deberían ser traducidos al español? ¿Por qué? ¿Hay excepciones?  
(Do you think all the traffic signs and other signs in public spaces in the U.S. should be translated into Spanish? Why? Are there exceptions?)