Revista de Lingüística y Lenguas Aplicadas
Vol. 17 año 2022, 59-70
EISSN 1886-6298
https://doi.org/10.4995/rlyla.2022.16513

THE EFFECT OF TASK MODALITY IN HERITAGE BILINGUALISM RESEARCH

Vicente Iranzo
Weber State University

Abstract: *Knowledge of the effects of task design features on linguistic performance is essential to obtain the right conclusions in linguistic research. Several studies have explored the effects of task modality in second language learners (Johnson, 1992; Murphy, 1997; Shiu, Yalçın, & Spada, 2018). These studies showed that L2 learners perform better when the stimulus is presented in written form compared to aural form. However, the effect of task modality in linguistic performance is an area that needs exploration in the field of heritage bilingualism research (Jegerski, 2018). Thus, the purpose of this study is to determine the effects of task modality on linguistic performance in heritage bilinguals. Participants in this study included 63 heritage bilinguals divided into high- and low-proficiency groups. All participants completed a proficiency test, a language background questionnaire, an aural speeded acceptability judgment task, and a written speeded acceptability judgment task. The statistical analysis revealed a significant effect of task modality on linguistic performance. High-proficiency heritage bilinguals performed significantly better (i.e., shorter reaction times) when the stimuli were presented in aural form. The results of this study contribute to the existing body on research methods in linguistic and specifically, to the effects that design features may have on linguistic performance in heritage bilinguals.*

Keywords: *heritage Spanish, research methods, task modality.*

1. INTRODUCTION

Previous research has underscored the importance of design features in linguistic research. For example, Murphy (1997) noted that the contradictory evidence regarding the role of Universal Grammar in the acquisition of L2 learners may lie in methodological issues, as these may lead to different results in performance. Along the same lines, Shiu, Yalçın, and Spada (2018) proposed that participants' performance on grammaticality judgment tasks —also known as acceptability judgment tasks— depends on learner-related factors, linguistic features, and task design variables. However, modality has received little attention compared to other task design variables (e.g., task stimulus). Therefore, Shiu et al. (2018) asserted that due to the prominent use of the acceptability judgment tasks in linguistic research, it is fundamental to explore how different task design variables, including modality, influence participants' performance.

The effect of task modality on participants´ linguistic performance has attracted the interest of several researchers in the Second Language Acquisition field. Previous research has shown that L2 learners obtain higher accuracies in acceptability judgment tasks when stimuli are presented in written form (instead of aural form). This better performance may be due to the fact that L2 learners tend to acquire the second language in formal setting when the exposure to written input is greater than in naturalistic acquisition. However, it has been proposed that heritage bilinguals generally have had more experience with spoken Spanish than with written Spanish (Jegerski, 2018). Hence, Jegerski (2018) proposed analyzing how linguistic performance varies depending on how the stimuli are presented to heritage bilinguals. Specifically, Jegerski (2018) stated that "given that heritage bilinguals typically have much more experience with oral language than with written text, then text-based measures might undershoot their Spanish ability" (p. 233). Therefore, the purpose of this study is to investigate how task modality affects linguistic performance in heritage bilinguals on speeded acceptability judgment tasks.

2. LITERATURE REVIEW

A limited number of studies have investigated the effects of the modality in L2 linguistic performance (Johnson, 1992; Murphy, 1997; Shiu et al., 2018). For example, Johnson (1992) used a written acceptability judgment task

To cite this article: Iranzo, V. (2022). "The effect of task modality in heritage bilingualism research". *Revista de Lingüística y Lenguas Aplicadas*, *17*, 59-70. https://doi.org/10.4995/rlyla.2022.16513
Correspondence author: vicenteiranzo@weber.edu

Received: 2021-10-20 Accepted: 2022-04-01

and employed the same materials as in Johnson and Newport's (1989) study. However, Johnson and Newport's (1989) study included an aural acceptability judgement task. The participants were native Chinese and Korean speakers learning English as a L2. Participants were undergraduate students, graduate students, faculty, and research associates at the University of Illinois, Champaign-Urbana. No English proficiency tests were included in this study, but all the participants had lived in the United States for at least five years. The written acceptability judgment task was designed to measure knowledge of English morphosyntax. Participants saw sentences written in English and had unlimited time to indicate whether they were grammatical or ungrammatical by circling the words 'yes' or 'no' next to the sentences. Twelve English rules were investigated, such as plural, past tense, and third person singular. The results of this task were compared to Johnson and Newport's (1982) study and indicated that there was substantially higher performance in the written task compared to the aural task; adult learners made 2.5 times more errors when the stimuli were presented aurally. The author proposed that this difficulty may have derived from "phonology, transitoriness of the stimuli, and the stimulus speed" (p. 243). Additionally, the author asserted that participants could use their explicit knowledge of English because of the untimed nature of the written task.

Murphy (1997) also compared the effects of modality on grammaticality judgment tasks. Eighty undergraduate and graduate students in Montreal were divided into four equally distributed groups: native English speakers, native French speakers, native English speakers learning French, and native French speakers learning English. L2 participants had learned their L2 after the age of 13 and were proficient in the L2 when the study was conducted. The English learners had an advanced level of English, according to the TOEFL sections administered. For the group of French learners, proficiency was measured indirectly. Both L2 groups took a cloze test in which they had to indicate the missing words in a paragraph in their non-native language (the French paragraph was the translation of the English paragraph). Since the L2 French learners performed significantly lower than the learners of English as a L2, the author concluded that their proficiency was different. The grammaticality judgment tasks, in English or French, contained 50 sentences: 20 practice sentences and 30 experimental sentences. Of the experimental sentences, 20 were target sentences and 10 were distractors. Also, there were two sets of sentences, and each participant saw only one version (ungrammatical or grammatical) of each pair. Each subject performed either the aural or written task, and accuracy results and reaction times were obtained. The results showed an impact of the modality, especially in the ungrammatical sentences, on the grammaticality judgment task performance because the participants performed higher when the stimuli were presented in a written modality. The researcher explained their results as follows: "Burdens of auditory processing seem to produce a greater obstacle for L2 learners to overcome than for native speakers" (p. 55). These results are in line with Johnson's (1992) study.

Shiu et al. (2018) also investigated the influence of the task modality in grammaticality judgement tasks. The participants were 181 native Mandarin-speaking university students from Taiwan who learned English as a L2. Also, 54 native English-speaking Canadians participated as a baseline comparison group. The Taiwanese students took the Oxford Placement Test as a proficiency measure. According to the European Reference Framework of Languages, the students obtained, on average, a B1 level (lower-intermediate level) in listening and a B2 (upper-intermediate level) in use of English. The study analyzed two structures: the be-passive voice and the past progressive. Four grammaticality judgement tasks were administered: two timed (aural-written) and two untimed (aural-written). In the timed aural/written tasks, the subjects listened to/read the entire sentence and were instructed to respond as quickly as possible. They could select among three answers: 'correct,' 'incorrect', and 'not sure.' In the untimed tasks, participants could listen to/read the sentences more than once. The results indicated that participants were less accurate when the stimuli were presented aurally rather than written, and the same trend was found when the results of timed and untimed tasks were compared. In addition, performance on the grammaticality judgment task was better on grammatical items than on agrammatical items. As for modality effects, the authors proposed that these results may be due to greater control in processing when items are presented in written form. Also, the researchers hypothesized that the effect in modality may have arisen due to the learners' language experience since these particular non-Western educated L2 learners tend to receive much more written than oral input. This argument coincided with the results obtained by the Oxford Placement Test, where the participants scored higher on the written than on the listening portion. These results are also aligned with studies conducted by Murphy (1997) and Johnson (1992).

Despite limited research, previous studies indicate that performance in acceptability judgment tasks improves in written tasks compared to aural tasks in L2 learners. However, further research is needed for several reasons. For example, each participant in Murphy's (1997) study performed the grammaticality judgment task in one modality (aural or written), and although the participants in Shiu, Yalçın, and Spada's 2018 study completed the tasks in both modalities, the order of the tasks was not counterbalanced. Therefore, task order arises as a confounding variable. Thus, to address the previous limitations, all participants in this study performed the tasks in both modalities, and the order of the tasks was counterbalanced.

While little is known about how modality affects linguistic performance in L2 learners (Shiu et al., 2018), these effects in heritage bilinguals are still relatively unexplored. Among unanswered questions that may direct future

research in the Heritage Acquisition field, Jegerski (2018) cautioned about the unknown effects of modality on processing. She considers that heritage bilinguals have generally received more aural than written input, so heritage bilingual performance may be weakened when obtained through written tasks in comparison to aural tasks. Therefore, there are still open questions regarding the effects of modality heritage bilingualism research. Consequently, it is a variable that should not be ignored and instead should be further explored.

3. THE PRESENT STUDY

The present study explored the following research question: is there a relationship of task modality (aural versus written) on the heritage bilingual's performance on a speeded acceptability judgment task as measured by accuracy and reaction time in speeded acceptability judgment tasks? Since heritage learners acquire the heritage language primarily in a naturalistic setting rather than through formal instruction like L2 learners, in this study it was hypothesized that heritage speakers would be more accurate and would have shorter reaction time in the aural task than in the written task.

4. METHODOLOGY

4.1 Participants

A total of 59 heritage bilinguals were recruited for the present study (four participants who were 2 Standard Deviations (SD) away from the mean of the proficiency group in either the aural or written modality were removed from the analysis). The heritage bilinguals who participated in this study were also enrolled in undergraduate or graduate classes. The average age of this group was 20.79 (*SD*=2.83). The university where this study was conducted has a specific Spanish program for heritage bilinguals. Therefore, as far as possible, Spanish course offerings try to ensure that heritage bilinguals continue their Spanish education in a program that meets their specific linguistic and socio-affective needs. However, the heritage program only offers second- and third-year courses, so heritage students who wish to complete a major also enroll in classes combined for them and L2 learners.

Participants completed the Spanish DELE proficiency test (see also Montrul & Bowles, 2009) using a computer. This test consists of 50 multiple-choice questions. Participants received one point for a correct answer and zero points for an incorrect answer. Depending on the score, participants were divided into two levels of proficiency: Low (0 to 35 points), and high (36 to 50). Table 1 shows the mean proficiency scores of the groups that participated in this study. This information is presented graphically in Figure 1.

Table 1. DELE Proficiency Test Results.

Proficiency	N	Mean	SD	SE
High	26	41.19	3.38	0.66
Low	33	23.52	7.73	1.34

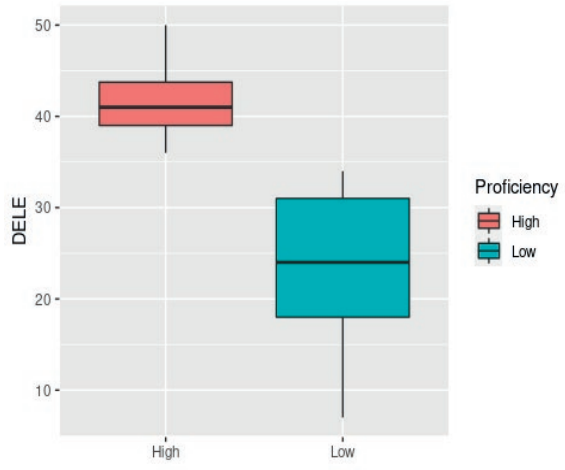

Figure 1. DELE Proficiency Test Results.

Also, participants completed the Bilingual Language Profile (Birdsong, Gertken, & Amengual, 2012). This questionnaire is free and can found in different language pairs and under a creative commons license. The questionnaire has a total of 19 questions using a Likert Scale and provides a numerical score of language dominance. The results of the questionnaire range from [-218,+218], where zero indicates balanced bilingualism. Positive scores indicate that the participant is dominant in English and negative scores indicate that the participant is dominant in Spanish. Table 2 includes the mean scores, standard deviations, and standard errors of the Bilingual Language Profile of participants according to proficiency. Figure 2 represents the scores of the BLP graphically.

Table 2. Bilingual Language Profile Results.

Proficiency	N	Mean	SD	SE
High	26	11.30	50.22	9.85
Low	33	63.89	44.25	7.70

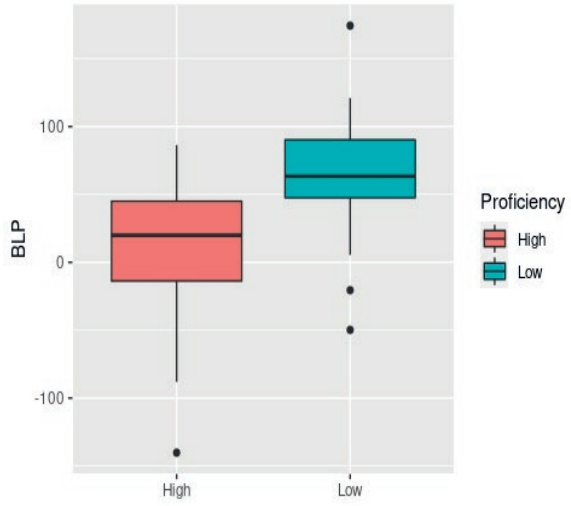

Figure 2. Bilingual Language Profile Results.

4.2 Written and Aural Speeded Acceptability judgment task

A crucial aspect of linguistic research is the dichotomy between implicit and explicit knowledge (Ellis, 2009). According to Spinner and Gass (2019), implicit knowledge is not conscious and, therefore, speakers are unaware that they have it. On the other hand, explicit knowledge is conscious, and speakers are aware of having it (Spinner & Gass, 2019). The tasks that may be used to promote the use of implicit knowledge are timed tasks in which the participants have a limited time to make a judgment (Spinner & Gass, 2019). Conversely, participants who perform untimed tasks can use the time they need to search their explicit knowledge to get the right answer (Spinner & Gass, 2019).

In the written speeded acceptability judgment task used in the present study, the participants had two seconds to give a judgment about the acceptability of a sentence. The target sentences included 12 pairs in which gender agreement in Spanish was analyzed. The target sentences contained seven to eight words. The target nouns that appeared in this task were singular and inanimate. The intention was to prevent participants from establishing gender agreement using biological gender cues (Foucart & Frenck-Mestre, 2012). Also, all nouns and adjectives followed the pattern of canonical grammatical gender in Spanish, i.e., masculine nouns end in *-o*, and feminine nouns end in *-a*. Masculine and feminine nouns were equally distributed.

The nouns and adjectives used in this task appeared in the book *Unidos: An interactive Approach* (Guzman, Lapuerta, & Liskin-Gasparro, 2019). This text is used in the first- and second-year Spanish courses in the institution at which the study was conducted. The intention was to choose appropriate vocabulary that was suitable for the low-proficiency students. In addition, once the list of words was created, a native speaker from the north of Mexico verified that the words belonged to the North Mexican linguistic variety. He found that the words *sabroso* ('tasty') and *marrón* ('brown') did not belong to his linguistic variety, so the words *rico* and *café* were used instead. Also, the task contained 56 filler sentences that included double objects, subject-verb agreement, or differential object marking. The fillers and the target items totaled 68, following the suggestion of Cowan and Hatasa (1995), who

recommended a maximum number of 72 sentences in order to reduce the fatigue of the participants (Spinner & Gass, 2019).

Two lists were created in the speeded acceptability judgment task, so the participants did not see both the grammatical and ungrammatical sentences of the same pair. For example, list 1 contained the grammatical option of pair 1 and the ungrammatical option of pair 2. On the contrary, list 2 contained the ungrammatical option of pair 1 and the grammatical option of pair 2, and so on. Once the two lists were created, the sentences were programmed in the Windows free software DMDX (Foster & Foster, 2003), which has been widely used in psycholinguistic studies for sentence, lexical, and semantic tasks (e.g., Arnon & Cohen Priva, 2013; Jiang & Nekrasova, 2007). DMDX allows for millisecond accuracy in both the presentation of aural and written stimuli and in the recording of responses (Forster & Forster, 2003).

The sentences were presented word by word with a stimulus-onset asynchrony of 500 ms. In addition, a Consolas fixed-width font in size 12 was used. Before starting the sentences, the participants saw a fixation point (+) for 500 ms. At the end of the sentence, the participants saw the question *¿Es aceptable la oración?* ('Is this sentence acceptable?') with the indications "Acceptable: left shift, not acceptable: right shift". This screen appeared for two seconds, the time allotted for participants to answer. After the two seconds, the message *Tiempo* ('Time') was shown and immediately the next sentence appeared.

Before starting the SAJTs, the participants had to familiarize themselves with the procedure by responding to three practice sentences. For each practice item, participants were given explicit feedback on whether or not the sentence was acceptable, the reason for such a judgment, and the key on the keyboard that had to be pressed in order to make the judgment. After the practice, the participants could ask the researcher any questions they had before the actual task started.

Once the lists were created with all sentences for the written task, a record of words contained in all the sentences was compiled and then sorted in alphabetical order. Subsequently, a native Spanish speaker born in northeastern Mexico audio recorded each word one by one. Recording words as a list was done in order to avoid providing prosodic cues to participants, these cues can inform the participant about the structure of the sentence (Marinis, 2010). When all the words were recorded, the lists used in the written task were programmed so that, instead of the written word, the audio of that word was played. At the end of the sentence, the procedure was identical to that for the written task. To clarify, the same items were used in both the written and aural tasks. The participants also completed a vocabulary and gender assignment task and a working memory test (the results of this task will be explored in another manuscript).

4.3 Procedure

The data for the present study were collected in two sessions. Figure 3 shows the order in which the instruments were administered over the two sessions.

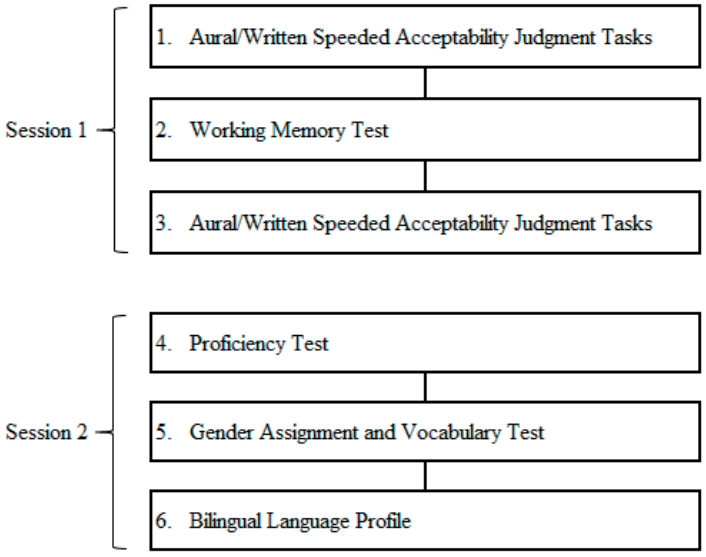

Figure 3. Order of Tasks Administered.

4.3.1 Session 1

The aural and written speeded acceptability judgment tasks and the WM test were performed at the University's language lab and research center. A PowerPoint presentation was shown, explaining the tasks with examples of the procedures. For the aural acceptability judgement task, the audio of all the computers was tested before the task to ensure participants were able to hear properly. After the task, they were asked if they had any problems with the quality of the audios. None reported difficulties with listening to the audios.

The participants were informed that the task contained three trial sentences before the actual task began and that at the end of the third sentence, they could ask any questions before continuing with the task. None of the trial sentences dealt with gender agreement in Spanish. The duration of each tasks was around 15 minutes. All participants were asked to remain silent until all classmates had concluded the tasks. The order of the aural and written tasks was counterbalanced. Three groups completed the aural task first while the other three completed the written task first. In all groups, a WM test was completed in between the two acceptability judgment tasks. Table 3 shows the order of the sessions by class.

Table 3. Order of AJTs.

Group	Acceptability Judgment Task 1	Acceptability Judgment Task 2
Group 1	Written	Aural
Group 2	Written	Aural
Group 3	Aural	Written
Group 4	Aural	Written
Group 5	Written	Aural
Group 6	Aural	Written

In addition, a group 7 comprising graduate students also participated in the study using the same materials as the undergraduate groups. However, this group did not perform the tasks during their classes but rather when the participants were available to voluntarily attend the language lab.

4.3.2 Session 2

During the second session, the participants first completed the proficiency test, followed by the vocabulary and assignment gender task, and finally, they completed the Bilingual Language Profile. For this session, the completion time was less than fifty minutes on average.

5. RESULTS

5.1 Descriptive statistics: response accuracy

The overall results (mean, standard deviation, and standard error) of all the gender agreement items in the acceptability judgment tasks are presented in Table 4. As observed in Table 4, there is a positive relationship between mean accuracy and level of proficiency in both groups. The mean accuracy for the high-proficiency participants was 74.20 (*SD*=16.29), and for the low-proficiency participants was 58.21 (*SD*=10.34). Figure 4 visually displays the results of mean accuracy and standard deviation, demonstrating the lower performance of low proficiency groups.

Table 4. Overall Mean Accuracy, SD, and SE in AJTs.

Proficiency	N	Mean	SD	SE
High	26	74.20	16.29	3.19
Low	33	58.21	10.34	1.80

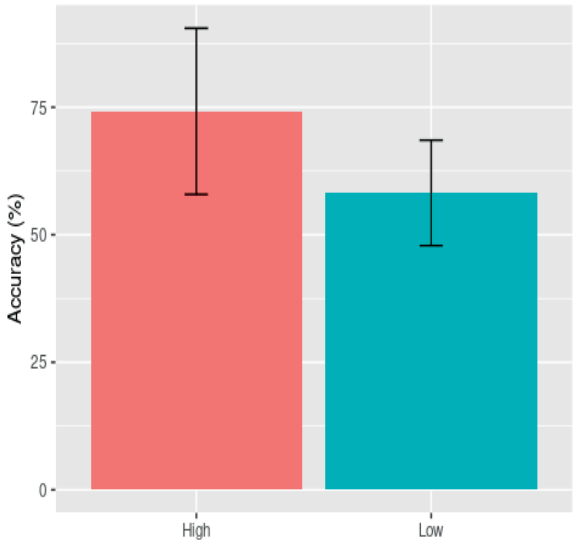

Figure 4. Overall Mean Accuracy and SD in AJTs.

In Table 5, the accuracy results of the written and aural modality are presented by group and proficiency level. Table 5 shows that both heritage bilingual groups obtained greater accuracy in the aural task than in the written task. The difference in accuracy was 4.01% for the high proficiency group and 1.47% for the low proficiency group. Figure 5 visually displays the results from Table 5.

Table 5. Mean Accuracy, SD, and SE in Different Modalities.

Proficiency	Modality	N	Mean	SD	SE
High	Aural	27	76.23	18.5	3.58
High	Written	27	72.22	17.6	3.39
Low	Aural	34	58.82	13.7	2.36
Low	Written	34	57.35	11.5	1.98

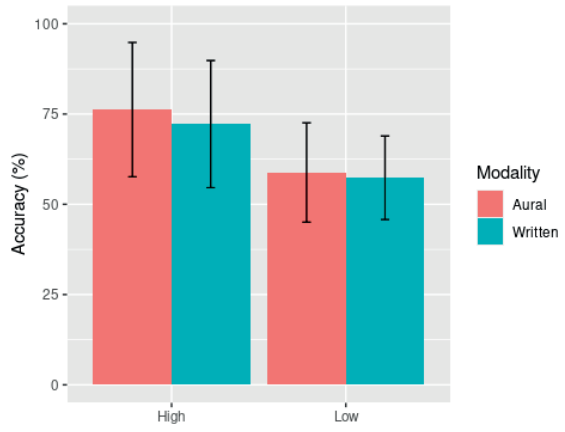

Figure 5. Mean Accuracy and SD in Different Modalities.

In Table 6, the accuracy results of the written and aural modality are presented by group, grammaticality and proficiency level. The high -proficiency heritage bilinguals obtained the best performance in the aural modality in the grammatical sentences (*M=74.36, SD=19.57*). On the contrary, the group of low-proficiency heritage bilinguals obtained the lowest accuracy, specifically in the aural modality and the agrammatical sentences (*M=41.41, SD=22.10*). Figure 6 visually displays the results of Table 6.

Table 6. Mean Accuracy, SD, and SE per Modality and Grammaticality

Proficiency	Modality	Grammaticality	N	Mean	SD	SE
High	Written	grammatical	26	74.36	19.57	3.84
High	Written	agrammatical	26	69.87	26.25	5.15
High	Aural	grammatical	26	75.00	23.21	4.55
High	Aural	agrammatical	26	77.56	25.80	5.06
Low	Written	grammatical	33	72.22	18.48	3.22
Low	Written	agrammatical	33	42.93	18.18	3.16
Low	Aural	grammatical	33	76.26	16.15	2.81
Low	Aural	agrammatical	33	41.41	22.10	3.85

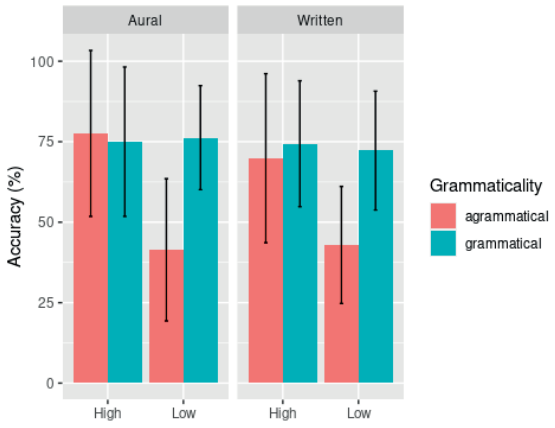

Figure 6. Mean Accuracy and SD per Modality and Grammaticality.

5.2 Descriptive Statistics: Reaction Times

Table 7 shows the mean reaction times, standard deviations, and standard errors of all groups by proficiency level. Figure 7 graphically shows the mean reaction times and standard deviations. The participants in the high-proficiency groups of heritage bilinguals showed shorter reaction times (*M*= 705.76, *SD*= 153.24), than did the low-proficiency group (*M*= 880.79, *SD*= 161.82).

Table 7. Overall Mean RT, SD, and SE in AJTs.

Proficiency	N	Mean	SD	SE
High	26	705.76	156.24	30.64
Low	33	880.79	161.82	28.17

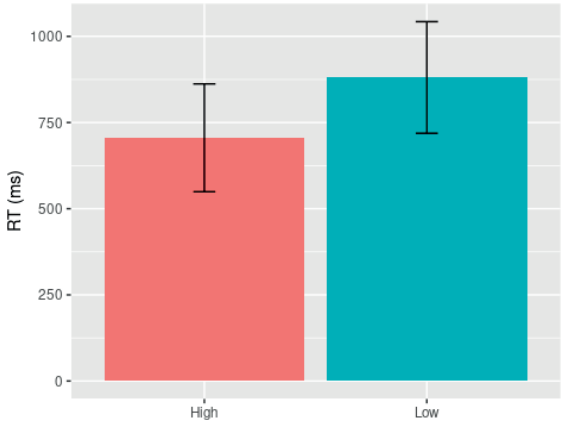

Figure 7. Overall Mean RT and SD in AJTs.

In Table 8, the reaction time results of the written and aural modality of the target items are presented. Both groups showed shorter reaction times when the stimuli was presented in aural form. The shortest reaction times were found in the two groups of participants in the oral modality. The difference was 82.93 ms for the high proficiency group and 60.8 ms for the low proficiency group. Figure 8 graphically provides the results of the mean reaction time and standard deviation in different modalities.

Table 8. Mean RT, SD, and SE in Different Modalities.

Proficiency	Modality	N	Mean	SD	SE
High	Aural	26	666.39	189.96	37.25
High	Written	26	749.32	194.09	38.06
Low	Aural	33	847.86	205.97	35.85
Low	Written	33	908.66	183.03	31.86

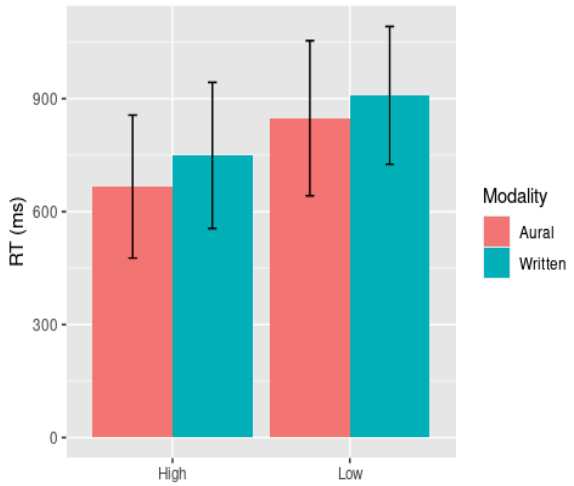

Figure 8. Mean RT and SD in Different Modalities.

In Table 9, the reaction time results of the written and aural modality are presented by group, grammaticality and proficiency. In both modalities, the high proficiency groups obtained shorter RT in the agrammatical sentences than in the grammatical sentences. However, the opposite situation occurred in the low proficiency group. Figure 9 graphically provides the results of the mean reaction time and standard deviation provided in Table 9.

Table 9. Mean Accuracy, SD, and SE per Modality and Grammaticality.

Proficiency	Modality	Grammaticality	Mean	SD	SE
High	Aural	agrammatical	588.01	242.81	48.56
High	Aural	grammatical	741.82	189.36	37.14
High	Written	agrammatical	746.19	299.51	58.74
High	Written	grammatical	785.74	276.62	54.25
Low	Aural	agrammatical	920.33	389.99	68.94
Low	Aural	grammatical	829.49	235.61	41.01
Low	Written	agrammatical	934.96	255.56	45.18
Low	Written	grammatical	888.84	228.22	39.73

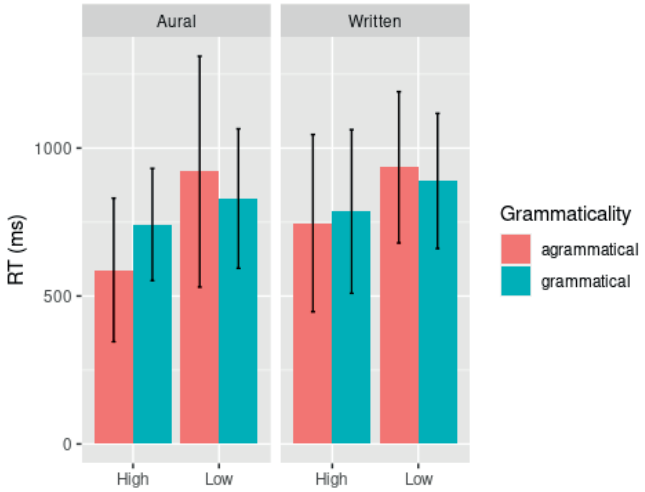

Figure 9. Mean RT and SD per Modality and Grammaticality.

5.3 Inferential Statistics

Both RT and accuracy data were analyzed through mixed-effects linear and logistic regressions in R version 3.6.3 (R Core Team, 2020). The models were fitted using the package lme4 v. 1.1.23 (Bates et al., 2015); after fitting, the estimated marginal means (EMM, Searle, Speed, and Milliken (1980)), contrasts and *p*-values were calculated using the package emmeans v. 1.5.0 (Lenth, 2020). The *p*-values reported here were corrected using Tukey's HSD test, as implemented in the latter package

All models were run with the maximal random effect structure as recommended in Barr et al. (Barr et al., 2013). The models were then incrementally simplified until they converged or were not a singular fit. For the RT models, the random structure included random intercepts by participant and item, and random slopes of modality by participant and item. For the accuracy models, the random structure only included random intercepts by participant ant item.

All models were compared against a null model (a variance component model with the same random structure as the model of interest). Both models were significantly different from the null model ($\chi^2(7) = 69.22$, $p < 0.01$ for accuracy and $\chi^2(7) = 39.69$, $p < 0.01$ for RT). Both were significantly different from the null model. Moreover, we visually assessed that the models met the assumptions of linearity with respect to the predicted value and of normality of residuals. Table 10 shows the model estimates for the variable accuracy and Table 11 shows the model estimates for the variable RT. No significant of modality was found for accuracy but a significant effect was found for RT (p=0.01). In this case, Modality interacts with grammaticality. Aural tasks lead to shorter RT on agrammatical sentences. Figure 10 and Figure 11 show the model estimates for accuracy and RT.

Table 10. Model Estimates for Accuracy.

Contrast	Gramm/Modality	Proficiency	odds.ratio	SE	df	z.ratio	*p*-value
Aural / Written	agrammatical	High	1.51	0.41	Inf	1.53	0.13
Aural / Written	grammatical	High	1.05	0.29	Inf	0.18	0.86
Aural / Written	agrammatical	Low	0.95	0.20	Inf	-0.25	0.80
Aural / Written	grammatical	Low	1.25	0.30	Inf	0.95	0.34
Agram / Gram	Aural	High	1.38	0.41	Inf	1.09	0.28
Agram / Gram	Written	High	0.96	0.28	Inf	-0.14	0.89
Agram / Gram	Aural	Low	0.24	0.06	Inf	-5.57	0.00
Agram / Gram	Written	Low	0.32	0.08	Inf	-4.57	0.00

Table 11. Model Estimates for RT.

Contrast	Gramm/Modal.	Profic.	estimate	SE	df	t.ratio	p.value
Aural / Written	agrammatical	High	-137.35	53.38	128.09	-2.57	0.01
Aural / Written	grammatical	High	-17.86	52.84	124.86	-0.34	0.74
Aural / Written	agrammatical	Low	-44.23	60.50	268.30	-0.73	0.47
Aural / Written	grammatical	Low	-66.55	47.34	115.93	-1.41	0.16
Agram / Gram	Aural	High	-178.81	51.15	198.84	-3.50	0.00
Agram / Gram	Written	High	-59.32	49.72	572.28	-1.19	0.23
Agram / Gram	Aural	Low	72.27	53.86	228.25	1.34	0.18
Agram / Gram	Written	Low	49.96	51.04	638.66	0.98	0.33

6. DISCUSSION AND CONCLUSION

This study investigated the effect of task modality as a task design feature in linguistic performance in heritage bilinguals. Although some heritage speakers have the opportunity to receive instruction in their heritage language in formal contexts (e.g. classroom instruction) where they are often exposed to the written modality of Spanish, heritage speakers tend to be exposed to the oral modality of Spanish. Therefore, this study aimed to explore which modality is advisable to use when conducting research with heritage bilinguals since the preference of one modality over another may be detrimental to the linguistic performance of heritage speakers.

Due to the acquisition patterns of heritage bilinguals, it was hypothesized that heritage bilinguals would be more accurate and have shorter reaction times in the aural AJT than in the written AJT. Although both groups of heritage bilinguals obtained a higher accuracy and responded faster in the aural modality, statistically significant differences were only found in the reaction times of the agrammatical sentences of the highly-proficient group of heritage bilinguals. In other words, modality effects were only found in high-proficiency bilinguals; these participants recognized errors significantly faster when the stimulus was presented in the aural modality.

Thus, this study found an effect of modality on task performance, supporting previous studies (Murphy, 1997, Johnson, 1992, Shiu et al., 2018); however, the favorable modality was different. While in previous research (Murphy,1997, Johnson,1992, Shiu et al., 2018), the better performance was found in the written task, this study found that heritage bilinguals performed better in the aural task. Nonetheless, there is a fundamental difference between this study and the studies by Murphy (1997), Johnson(1992), and Shiu et al. (2018) that may explain these results. Whereas the aforementioned studies included L2 learners as participants, the present study included heritage bilinguals. As previously stated, heritage bilinguals acquire their heritage language in a naturalistic environment with potentially reduced opportunities for formal instruction in their heritage language. That is, heritage speakers tend to receive more aural than written input than L2 learners. In this line, Shiu, Yalcin, and Spada's (2018) concluded that, in the case of L2 learners, the poorer performance in the aural task might be due to the formal learning conditions where learners tend to receive more written input than aural input. In this case, we have the opposite situation; the poorer performance in the written modality in heritage speakers is that they tend to receive more aural input than written input.

Also, this study is not without its limitations. For the effects of modality on the task for acceptability judgment, it is worth noting that all heritage bilinguals were enrolled in academic Spanish language, literature or culture classes at the time of the study. The participant pool and research question would have been greatly enriched had we been able to recruit heritage bilinguals not enrolled in academic Spanish classes.

To conclude, the present study has contributed to the area of linguistic research methods by evaluating the impact of modality as a design feature. This study found that the impact of modality on linguistic performance affects heritage bilinguals, specifically high-proficiency learners, where better performance was observed in the aural modality. Thus, the performance of the participants in an AJT does not depend solely on the structure being analyzed. Task design features can affect performance.

REFERENCES

Arnon, I., & Cohen Priva, U. (2013). "More than words: The effect of multi-word frequency and constituency on phonetic duration". *Language and Speech*, *56*/3: 349–373. https://doi.org/10.1177/0023830913484891

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". *Journal of memory and language, 68*/3: 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software, 67*/1: 1–48. https://doi.org/10.18637/jss.v067.i01

Birdsong, D., Gertken, L.M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to use instrument to assess bilingualism*. Center for Open Educational Resources and Language Learning, University of Texas at Austin. Retrieved from https://sites.la.utexas.edu/bilingual/

Cowan, R., & Hatasa, Y.A. (1994). "Investigating the validity and reliability of native speaker and second-language learner judgments about sentences". In E. Tarone, S. M. Gass, & A. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 287–302). Hillsdale, NJ: Lawrence Erlbaum.

Ellis, R. (2009). "Implicit and explicit learning, knowledge and instruction". In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691767-003

Forster, K., & Forster, J.C. (2003). "DMDX: A Windows display program with millisecond accuracy". *Behavior Research Methods, Instruments and Computers*, *35*/1: 116–124. https://doi.org/10.3758/BF03195503

Foucart, A., & Frenck-Mestre, C. (2012). "Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking". *Journal of Memory and Language*, *66*/1: 226–248. https://doi.org/10.1016/j.jml.2011.07.007

Guzmán, E.E., Lapuerta, P., & Liskin–Gasparro, J.E. (2019). *Unidos: An interactive approach*. New York: NY: Pearson.

Jegerski, J. (2018). "Psycholinguistic perspectives on Spanish as a heritage language". In K. Potowski (Ed.), *Routledge handbook of Spanish as a heritage/minority language* (pp. 221–234). New York, NY: Routledge. https://doi.org/10.4324/9781315735139-15

Jiang, N., & Nekrasova, T.M. (2007). "The processing of formulaic sequences by second language speakers". *The Modern Language Journal*, *91*/3: 433–445. https://doi.org/10.1111/j.1540-4781.2007.00589.x

Johnson, J.S. (1992). "Critical period effects in second language acquisition: The effects of written versus auditory materials on the assessment of grammatical competence". *Language Learning*, *42*/2: 217–248. https://doi.org/10.1111/j.1467-1770.1992.tb00708.x

Johnson, J.S., & Newport, E. (1989). "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language". *Cognitive Psychology*, *21*/1: 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Lenth, R. (2020). Emmeans: Estimated Marginal Means, Aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans.

Marinis, T. (2010). "Using on-line processing methods in language acquisition research". In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1075/lllt.27.09mar

Montrul, S., & Bowles, M. (2009). "Back to basics: Differential object marking under incomplete acquisition in Spanish heritage speakers". *Bilingualism: Language and Cognition*, *12*/3: 363–383. https://doi.org/10.1017/S1366728909990071

Murphy, V. (1997). "The effect of modality on a grammaticality judgment task". *Second Language Research*, *13*/1: 34–65. https://doi.org/10.1191/026765897671676818

R Core Team (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Searle, S.R., Speed, F.M., & Milliken, G.A. (1980). "Population marginal means in the linear model: an alternative to least squares means". *The American Statistician, 34*/4: 216-221. https://doi.org/10.2307/2684063

Shiu, L-J., Yalçın, Ş., & Spada, N. (2018). "Exploring second language learners' grammaticality judgment performance in relation to task design features". *System*, *72*, 215–225. https://doi.org/10.1016/j.system.2017.12.004

Spinner, P., & Gass, S.M. (2019). *Using judgments in second language acquisition research*. New York, NY: Routledge. https://doi.org/10.4324/9781315463377