Applied Psycholinguistics (2026), **47**, e45, 1–25
doi:10.1017/S0142716426100873

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Variable bilingual experiences and their influence on a contextualized acceptability judgment task

Rodrigo Delgado[1], Salvatore Callesano[1] and Wenyue Ma[2]

[1]Department of Spanish and Portuguese, University of Illinois Urbana-Champaign, USA and [2]School of Literatures, Cultures, & Linguistics, University of Illinois Urbana-Champaign, USA
Corresponding author: Rodrigo Delgado; Email: rdelga21@illinois.edu

(Received 4 August 2025 UTC; revised 27 May 2026 UTC; accepted 3 June 2026 UTC)

Abstract
This study examines the relationship between code-switching (CS) experience, as measured by the Bilingual Code-Switching Profile (BCSP), and ratings from an acceptability judgment task (AJT). Although the BCSP offers a structured way to quantify bilinguals' CS behavior, its application in grammaticality studies and its ability to reflect bilingual heterogeneity have been underexamined. We address this by asking: (1) How do BCSP scores predict acceptability ratings of Spanish–English CS sentences? and (2) How do BCSP scores reflect the social heterogeneity of US bilingual populations, focusing on age, gender, region, place of birth, and national origin? A total of 137 Latinx bilinguals across the USA completed an online survey including the BCSP and 40 reggaetón-inspired CS sentences, balanced for grammaticality and experimental label (i.e., Source: Reggaetón lyrics). Results from multilevel and linear regression models indicate that higher BCSP scores are associated with higher AJT ratings, particularly among participants who report frequent CS use, greater proficiency, and positive attitudes. Social variables had limited impact, although age and place of birth emerged as modest predictors. These findings highlight the BCSP's value in capturing CS experience and stress the need to account for bilingual variability in studies of acceptability judgments.

Keywords: Acceptability judgment task (AJT); Bilingual Code-Switching Profile (BCSP); code-switching; Spanish–English bilinguals; variation

Introduction

Code-switching (CS) is a heterogeneous practice, with no two bilinguals having the same linguistic behavior. This variation arises from several factors, such as age of acquisition, community norms, etc. Consequently, results of grammatical acceptability of CS may be less reliable if individual linguistic backgrounds are not systematically controlled (Gonzalez-Vilbazo et al., 2013). A systematic, reliable, and

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

valid tool for measuring CS behavior would go a long way in addressing this potential issue. In this article, we focus on one such instrument: the Bilingual Code-Switching Profile (BCSP, Olson, 2022).

The BCSP is a valuable tool for advancing the way researchers methodologically capture real-world CS behavior. Moreover, the tool has been applied to research on acceptability judgments (Olson, 2025); however, additional applications of the BCSP to acceptability judgment research are limited. Furthermore, the extent to which social factors affect BCSP scores remains largely unexplored. Given that bilingual speakers do not engage in CS uniformly, and that various factors, such as participant demographics and experiences with CS, vary widely across individuals and populations, the need to empirically investigate the relationship between social factors, BCSP scores, and acceptability judgments becomes essential.

In this article, we fill these gaps with two main objectives: first, to examine if acceptability of Spanish–English CS sentences is conditioned by BCSP scores, and second, to test whether BCSP scores are conditioned by social variables, such as age, gender, region of residence, place of birth, and family national origin. Although our current approach is inspired by a usage-based framework often used in sociolinguistic variation research (e.g., Backus, 2015), our article directly contributes to contemporary psycholinguistic discussions on context and heterogeneity within bilingual populations (e.g., López et al., 2023). As Sprouse (2018) notes, when it comes to language external factors that may influence acceptability judgment task (AJT) results, most research tends to focus on methodological concerns, such as repetition effects, statistical power, and naïve vs. expert participants. This last point notes some variation within the participant populations, and it is to that topic that we make an experimental expansion in our research. As such, one primary aim of this work is to illustrate how context, both in terms of experimental stimuli and social considerations, is critical to the processing of grammaticality, especially in mixed-language settings. In this way, the research outlined in the following sections provides context for a psycholinguistic test, the AJT, while also focusing on variable experiences within the bilingual populations we study.

By integrating the BCSP into an experimental design with culturally relevant stimuli, specifically reggaetón[1]-inspired CS sentences, this study contributes to the growing body of research examining the role of lived experiences in shaping linguistic behavior. Importantly, we link individual CS experience to linguistic processing, as measured through acceptability judgments, which helps bridge the gap between historically separate disciplines and illustrates the importance of focusing on context in bilingual acceptability judgments.

Literature review

The following sections outline the three threads of research that motivate the present study: (i) methods to measure CS behavior; (ii) experimental work using AJTs to study CS; and (iii) psycholinguistic research highlighting heterogeneity within bilingual populations. Together, these three threads underscore the importance of considering measures of CS engagement and experience in experimental CS research.

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

The BCSP

Many factors affect how and when a bilingual engages in CS; a bilingual typically adheres to grammatical constraints (e.g., Macswan, 2005; Myers-Scotton, 1993; Poplack, 1980), to social contexts (e.g., Zentella, 1997), and attitudes and biases (Anderson & Toribio, 2007; Badiola et al., 2018; Rangel et al., 2015; Regan & Toribio, 2026). However, although researchers have identified a range of factors that influence CS, few studies have systematically measured CS behavior—a gap that Olson (2022) addresses with the BCSP. Olson (2023) then reviewed empirical studies focusing on CS, analyzing 140 articles from a wide range of linguistic journals whose keywords included *bilingualism* and *code-switching*, among others. Although the original study focused on how proficiency is defined and measured in bilingualism research, a reanalysis of these studies showed that only 2.9% ($n = 4$) measured CS behaviors. Furthermore, Olson (2022) states that there is currently no consensus on how to measure CS practices, even among those that do measure it, whether directly or indirectly. An example of a direct measure, on the one hand, is calculating the number of CS by length of conversation in semistructured interviews (Yim & Bialystock, 2012). Indirect measures, on the other hand, can take the form of self-reports, indicating how often, when, where, and with whom one code-switches. However, although these self-reports measure some aspects of CS, such as frequency, they do not account for key variables that potentially shape CS behavior, such as a bilingual's engagement with CS long term (CS history), current or recent use in a variety of contexts (CS use), self-reported proficiency within CS speaking, listening, reading, and writing (CS proficiency), and attitudes toward CS (CS attitudes) (Olson, 2022). The BCSP, then, moves the field one step closer to having a standardized, detailed, and systematic way of assessing bilingual experiences and CS practices and behavior (Olson, 2022).

For evidence of the test–retest reliability and validity of the BCSP, see Olson (2022), who argues that the global BCSP score, along with each individual subscore, provides strong reliability. However, Olson does provide two limitations, which are the points of departure for our current analyses. First, Olson (2022) states that although the internal validity of the BCSP has been supported, "future research on the BCSP may seek to further establish the external validity of the instrument correlating the self-reported BCSP scores with a variety of behavioral measure of CS practices, including observational measure of CS usage [and] grammaticality judgments" (p. 424). Second, Olson states that, although the BCSP was validated for a specific population, future work "should extend beyond this limited population ... expanding on the ... diversity of the target population" (p. 423). In the present article, we address these two limitations by correlating BCSP scores with CS judgments and further examining if factors such as age, gender, region of residence, place of birth, and family national origin, condition BCSP results from a national sample.

AJTs in research

The AJT is a frequently used experimental task in CS research, which has become a valuable tool because it allows us to "access to combinations of linguistic features

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

that may otherwise be difficult (or impossible) to observe in monolingual data" (González-Vilbazo et al., 2013; p. 2). Moreover, we can study CS structures that are not common in natural oral production, such as the mixed DPs where the determiner is in English and the noun in Spanish (e.g., the *mesa* "the table") (Herring et al., 2010) or subject pronoun switches (*ellos* run "they run" vs. *esoschicos* run "those guys run") (González-Vilbazo & Koronkiewicz, 2016). In AJTs, participants are provided with sentences to judge on a, usually, 7-point scale, ranging from *completely unacceptable (1)* to *completely acceptable (7)*. AJTs have been used to explore the possible constraints of human language by testing whether a sentence is part of a speaker's grammar. The effectiveness of this task has been supported by its replicability (Sprouse & Almeida, 2017) and ecological validity (Hofweber et al., 2019), cementing it as an invaluable tool that has contributed to our knowledge of syntactic theory. Although AJTs have historically been a monolingual experimental task, their use has been extended to bilingual CS research (e.g., González-Vilbazo et al., 2013), albeit with some drawbacks. A possible limitation of AJTs in CS research is the decontextualization of the CS utterances. For example, previous literature has shown that bilinguals not only adhere to structural constraints when CS, but also to sociolinguistic constraints (e.g., Zentella, 1997). This has led to an important line of research examining the factors that could affect AJT ratings, shifting the AJT to be a useful tool in usage-based approaches to language (Backus, 2015).

In this article, we argue that social forces influence language contact and mixing (i.e., CS) (Parafita Couto et al., 2021; Rojas-Berscia, 2024), and that social forces, such as region of residence, have an impact on AJT ratings. Balam et al. (2020), for example, demonstrated that the use and acceptability of CS compound verbs, composed of the Spanish light verb *hacer* "to do" and an English verb in the infinitive (e.g., *hicieron* rent "they rented"), are conditioned by community norms. They tested speakers from Northern Belize, New Mexico, and Puerto Rico and found that the speakers from Northern Belize gave the highest ratings for the *hacer* construction, although US bilinguals preferred the *estar* construction (*estar* training "to be training"). In another experiment, Blokzijl et al. (2017) tested CS determiner phrases (e.g., *la* table vs. the *mesa*"the table") in Spanish–English bilinguals in Miami and Spanish–English Creole bilinguals in Nicaragua. They concluded that these two populations behaved differently regarding the choice of language for the determiner, where Miami bilinguals preferred the Spanish determiner (e.g., *la* table "the table"), Nicaraguan bilinguals preferred the Nicaraguan Creole English determiner, *ur*, and the Spanish noun, *luna* (e.g., *ur luna* "the moon"). Indeed, as Parafita Cuoto et al. (2021) conclude, many asymmetries that are found in CS research "are due to extralinguistic factors, such as community norms, rather than structural properties of the participating languages" (p. 17).

Besides community norms, studies have shown that variability in CS judgments are shaped by sociolinguistic context and experiential factors. One key dimension is attitudes toward CS. For example, in the study of Badiola et al. (2018), 50 Spanish–English early bilinguals answered a questionnaire to assess their attitudes toward CS and completed a CSAJT. The authors found that positive attitudes toward CS resulted in higher acceptability ratings. In comparison, negative attitudes toward CS resulted in lower ratings, showing that language attitudes are a confounding

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

factor that must be considered. Another key factor is sociolinguistic experience with CS. For example, Callesano et al. (2025) report that higher listening frequency of reggaetón resulted in statistically significantly higher AJT ratings. The authors highlight the impact of sociolinguistic context and experience on CS acceptability judgments.

Previous research (Balam & Shelton, 2023; Callesano et al., 2025; Naranjo Hayes, 2022) has explored a sociocultural context specifically relevant to CS, namely reggaetón. Monteagudo (2020) noted that CS in reggaetón was more frequent in 2018 than in earlier years, concluding that it has become almost a necessity for a song's success. It should be of no surprise, then, that CS "data from songs can align with patterns attested in conversational [CS] in U.S. Spanish/English bilingual communities" (Balam & Shelton, 2023, p. 47). This is further supported by Callesano et al. (2026), who found that when Spanish–English bilinguals are told CS sentences come from reggaetón lyrics, their acceptability increased, but this variation was conditioned by the participants' place of birth and state of residence. Although reggaetón-source labels are not the primary focus on the current analyses, they constitute an important methodological manipulation. Furthermore, we highlight the added benefit that we can test CS sentences in a culturally relevant context, allowing us to examine how experience, attitudes, and exposure affect judgments.

Ultimately, bilingual acceptability judgments are conditioned not only by grammatical constraints but also by sociolinguistic variables. This raises the question of how individual differences should be controlled in CS experimental designs.

Bilingual heterogeneity

Within the fields of psycholinguistics and cognitive science, there has been a recent surge in interest to focus on variation within bilingual populations and bilingual experiences to produce more inclusive, holistic, and interdisciplinary research (López et al., 2023). This call for more fine-grained analyses, without monolingual comparisons, for example, has led researchers to focus more on individual differences. For example, Beatty-Martínez and Dussias (2017) tested code-switchers and non-code-switchers in a series of three experiments testing determiner–noun switches, as well as eliciting naturally produced CS speech. They found that CS cost depended on the type of CS and the bilingual's CS experience. López et al. (2019) highlighted how language brokers engage in informal translation practices across a range of domains, and these variable experiences lead to differential attitudes toward bilingualism and correlate with an increase in problem-solving abilities. When testing the processing of figurative language (i.e. idioms), López and Vaid (2018) demonstrated that, when testing words in Spanish, the accuracy scores for language brokers were significantly higher than non-brokers. These examples also corroborate other findings that highlight heterogeneous experiences of bi/multilingual populations. Previous research has demonstrated that personality mediates CS production (Dewaele & Li, 2014), cognitive effects are conditioned by age of acquisition and proficiency ((Bonfieni et al., 2019), geographic region illustrates variation of bilingual verb switches (Balam et al., 2020), and interactional context

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

significantly impacts how bilinguals access and control their languages (Green & Abutalebi, 2013).

Taken together, and principally inspired by López et al. (2023), considering the inherent variation and heterogeneity of bilingual histories and experiences is critical for achieving more holistic models of bilingualism. This approach is in alignment with usage-based theories of grammar, which suggest that linguistic representation is influenced by frequency and context of use (Bybee, 2006). Although our current work relies on offline acceptability judgments rather than online processing measures such as reading times or Event Related Potentials (ERPs), we argue that by understanding how heterogeneous bilingual experiences influences CS acceptability judgments, we add to ongoing discussions that focus on individual variation in cognitive science and psycholinguistics.

The present study integrates these insights by using the BCSP as a tool to quantify bilinguals' real-world CS engagement. Then, by relating BCSP scores to judgments on an AJT, we explore how experience with CS may shape the cognitive decisions toward bilingual grammaticality.

Research questions

In this article, we aim to expand our knowledge of the extralinguistic factors that affect AJTs. To expand on research connecting the BCSP with AJTs and to further understand the variable bilingual experiences captured by the BCSP, we present the following research questions:

RQ1: How are AJT results of Spanish–English CS sentences conditioned by previous experiences with CS, as understood through the BCSP scores?

RQ2: How are the results of the BCSP conditioned by social variables?

Materials and methods

Participants

In collaboration with CloudResearch panels between December 2023 through February 2024, a total of 137 participants were ultimately included in our study. All participants self-identified as Hispanic/Latinx. The average age was 36.3 years (SD = 13.8, range = 18–78). The sample included 48 males and 89 females. Originally, the researchers chose five major cities that had a large Spanish-speaking population. However, participant recruitment at this level was not feasible, and research was extended to the states. Participants were drawn from five US states: California (n = 26), Florida (n = 28), Illinois (n = 27), New York (n = 26), and Texas (n = 30). The participants were asked, as an open-ended question, to write-in their family national origin. The researchers then grouped the responses, as balanced as possible, into categories without eliminating participants or erasing their self-reported demographics. The following were the groupings: 66 identified as Mexican, 40 as Caribbean, and 27 as South or Central American, with 4 missing

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

responses excluded from the analysis. Regarding place of birth, 87 participants were born in the USA and 50 were born outside the USA.

Design

All participants completed an online survey created with Qualtrics. The survey was written entirely in Spanish–English CS, and it was sent to three Latinx bilingual consultants, two in Chicago and one in El Paso. They were asked to read the survey and highlight any element of the survey that they deemed to be unacceptable, which were then modified. The survey consisted of four sections. In the first section, participants answered two screening questions to confirm that only bilingual code-switchers were included in the sample. The first questions asked: "Are you a bilingual (English–Spanish)?" The participant had the option to click on *"sí* 'yes'" or "no." If they clicked "no," they were not allowed to continue with the survey. The second question asked: "What is the word *con la que empieza esta pregunta?* 'What is the word with which this question starts?'" The participants had the option to click on: (i) word, (ii) *empieza*, (iii) what. Any other option besides "(iii) what" also ended the survey. These two questions were developed specifically for this study by the researchers. In the second section, the participants consented to participate in the study voluntarily.

In the third section, participants were shown 40 Spanish–English code-switched sentences based on a 2 × 2 design. The stimuli varied along two dimensions: grammaticality (grammatical vs. ungrammatical) and the presence of a reggaetón label (with vs. without). Drawing on prior studies that incorporate extralinguistic cues in experimental materials (e.g., Callesano & Carter, 2019; Carter & Callesano, 2018; Hay & Drager, 2010; Niedzielski, 1999), ostensible source labels were included for half of both the grammatical and ungrammatical items. The purpose of the source label was to tell the participant that the CS item they were rating originated from a reggaetón song. Each of the four conditions—grammatical with label, grammatical without label, ungrammatical with label, and ungrammatical without label—included 10 items. Two examples of how the items were presented to the participants (one grammatical without label and the other ungrammatical with label) are provided in Figures 1 and 2 in the Appendix. All the sentences were inspired by reggaetón lyrics, specifically from three songs (Fue Mejor by Kali Uchis and SZA, //aguardiente y limón %◡‿‿◡% by Kali Uchis, and The Queen is Here by Ivy Queen), and modifications were made to remove vulgar content. Participants were not made aware of any specific artists, and no item in our survey is a direct copy of an actual song lyric. Half of the items were changed manually to be ungrammatical, following Koronkiewicz (2014), by using pronominal subject switches instead of full noun phrase subjects (e.g., **Ella* went to the party "she . . . "). Across all 40 items, the mean sentence length is 6.85 words (7.55 words for the grammatical category and 6.15 words for the ungrammatical category). The stimuli sentences were not lexically identical due to methodological concerns. In the current design, if they were presented with sentences that only differed in the pronominal subject switches, the participants would have become aware of the source label experimental manipulation. In the present study, having the stimuli vary lexically, while controlling the CS structure, allows us to preserve the plausibility that the

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

stimuli were reggaetón-inspired, eliciting more naturalistic judgments. Furthermore, the researchers kept the stimuli as semantically similar as possible (i.e., as possible reggaetón lyrics) to avoid this being a confounding variable. Furthermore, our design considers the concept of "reggaetón" with a broad definition; however, as noted by Naranjo Hayes (2022), one of the ways in which different artists vary in their CS production is switch direction. Although we do not consider this variable in our analyses, to avoid over complicating our survey and statistical designs, we present a description of the switch directions in our survey items. As the ungrammatical items were manually manipulated to switch at the boundary between the subject pronoun and the verb, this category includes 10 switches in each direction (Spanish to English and English to Spanish). The grammatical switches, whose syntax was not manually manipulated, include more variation in direction. There are 14 instances of English to Spanish switches, and 24 instances of Spanish to English switches. Some items (e.g., *Me dice* everything and nothing *al mismo tiempo*) include switches in both directions.

The CS stimuli were presented in a randomized order for each participant, with each sentence followed by four questions: (1) how natural the sentence sounds? (2) how likely it is to be heard in the participant's community? (3) how likely the participant is to use it themselves? and (4) how strongly the participant believes the sentence is correct? All responses were recorded on a 7-point likert scale, where higher scores indicated greater likelihood, agreement, or perceived acceptability. The fourth and final section of the survey gathered demographic information.

Participants completed the BCSP in the third part of the study. The BCSP consists of four main sections—CS use, CS attitudes, CS proficiency, and CS history—modeled after the Bilingual Language Profile (BLP; Birdsong et al., 2012). Although the BCSP had previously been available for various language pairs, a version written in CS had not yet been developed. Using the existing English and Spanish versions of the BCSP as a guide, the first two authors created a code-switched version for this study. All materials of the current research are available on the Open Science Framework (OSF) page accompanying (https://osf.io/6gcxp/).

General analytical methods

In this section we outline the analytical procedures used to operationalize our variables and to model the relationships between CS experience and acceptability judgments. For the BCSP survey, we followed Olson (2022) and calculated a total score from the participants' scores across the four sections. Each section's score is calculated based on the responses to its respective questions. To ensure that each section contributes equally to the overall score, the raw scores for each section are multiplied by a specific weighting coefficient. The weighting coefficients are designed to standardize the contribution of each section to the total score. The coefficients ensure that each section (CS history, CS use, CS proficiency, and CS attitudes) accounts for 25% of the total score. Specifically, the coefficients are as follows: CS history is weighted by 0.2083, CS use by 0.6944, and both CS proficiency and CS attitudes by 1.0417. The weighted scores from each section are then summed to produce the final BCSP score, which ranges from 0 to 100. In this continuous

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

measure, lower scores indicate less engagement and experience with CS, although higher scores indicate greater engagement and experience with CS.

For the AJT results, we computed a composite acceptability score by averaging the ratings across the four questions listed in the Design section. This approach was supported by strong inter-item correlations (ranging from 0.79 to 0.84), indicating a high degree of internal consistency. Additionally, this composite score demonstrated strong correlations with each individual score ($r = 0.93$–0.94), further validating its use as a representative measure of acceptability. A principal component analysis (PCA) confirmed the suitability of this measure, with the Kaiser rule and parallel analysis suggesting a single dominant component (Eigen values: 3.42, 0.22, 0.21, 0.16). The PCA-derived component score correlated almost perfectly with the mean score ($r = 0.9998$), reinforcing the reliability of the composite measure. Consequently, we used this composite AJT score in our subsequent statistical analyses.

For the first research question, to examine how AJT results of Spanish–English CS sentences are conditioned by previous experiences with CS, as measured by BCSP scores, we employed multilevel mixed-effects regression models using the lme4 package in *R*. This approach accounts for the nested structure of the data, where responses are grouped within participants and items, allowing us to model participant- and item-specific variabilities. We estimated *p*-values for the fixed effects using Satterthwaite's approximations and calculated confidence intervals via the Wald method to ensure reliable interval estimates.

For the mixed-effects models addressing RQ1, the fixed-effects structure was specified a priori based on the research questions and the 2 × 2 experimental design. Each model included the relevant BCSP measure, grammaticality, label status, and their interactions. These interaction terms were retained because they directly tested whether the relationship between CS experience and AJT ratings varied as a function of grammaticality and/or reggaetón label status. For this reason, we retained the full fixed-effects structure across models rather than using a data-driven stepwise selection procedure. Model comparison procedures were instead used to determine the most appropriate random-effects structure. We began with a random-intercept model including random intercepts for participants and items, then added theoretically justified random slopes when supported by the data. Competing random-effects structures were compared using likelihood-ratio tests, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The final random-effects structure was selected based on converging evidence from model comparisons and information criteria, with the goal of improving model fit while avoiding unnecessary complexity.

Before turning to the mixed-effects modeling results, we first present descriptive statistics to provide an overview of acceptability patterns across the experimental conditions. Table 1 reports mean composite AJT scores and standard deviations for each condition in the 2 × 2 design of grammaticality by label. These descriptive results offer an initial view of the data and help situate the mixed-effects modeling that follows.

Building on these descriptive patterns, we then fit mixed-effects regression models to examine the factors shaping AJT judgments. The dependent variable in our models was the composite AJT score, although the independent variables

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 1. Descriptive statistics for AJT ratings by grammaticality and label status

By grammaticality	Mean	SD	95% CI
Ungrammatical	4.17	1.96	[4.10, 4.24]
Grammatical	4.57	1.83	[4.50, 4.64]
By label status	**Mean**	**SD**	**95% CI**
Without	4.36	1.93	[4.28, 4.43]
With	4.39	1.88	[4.32, 4.46]

included participants' BCSP scores, item grammaticality, and reggaetón label status. We allowed for all possible interaction effects among these factors. To gain a comprehensive understanding of how different aspects of CS experience influence AJT judgments, we constructed five separate mixed-effects models: one using the total BCSP score as the predictor and four additional models examining the effects of each BCSP section score (CS history, CS use, CS proficiency, and CS attitudes). This approach allowed us to assess both the overall influence of prior CS experience and the distinct contributions of its individual components. To evaluate the robustness of the fixed-effects specification, we also conducted a sensitivity analysis in which each full interaction model was compared with a corresponding main-effects-only model. This comparison was used to assess whether the inclusion of theoretically motivated interaction terms affected the substantive interpretation of the findings.

In the mixed-effects model output, we report marginal R^2 and conditional R^2 as model fit indices. Marginal R^2 reflects the variance explained by fixed effects, highlighting the explanatory power of key predictors. Conditional R^2 accounts for both fixed and random effects, capturing the overall model fit, including individual and item-level variability. Higher marginal R^2 suggests that fixed effects explain a substantial portion of the variance, although higher conditional R^2 indicates strong model performance, incorporating random effects.

Before estimation, all predictors were re-coded to minimize multicollinearity and to make the coefficients directly interpretable. Each BCSP measure–both the total score and each of the four section scores–was grand-mean centered, retaining its original scale while redefining zero as the sample mean; the model intercept therefore represents the predicted AJT rating for a participant with a BCSP score averaged across grammatical and ungrammatical (and across labeled and unlabeled) sentences. Grammaticality and label status were effect-coded (−0.5, +0.5), with +0.5 representing grammatical sentences and sentences with a reggaetón label, and −0.5 representing ungrammatical sentences and sentences without a label. Under this coding, the coefficients represent the mean difference between the two levels of each factor. Without this centering and symmetric coding, the cross-level interaction term (BCSP × Grammaticality) was almost perfectly correlated with its parent variables, inflating standard errors and obscuring significance tests. The revised coding rendered all variance-inflation factors <2, removed convergence warnings,

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

and yielded fixed-effect estimates whose magnitude and precision can be interpreted with confidence.

For the second research question, to examine how BCSP scores are conditioned by demographic factors, we conducted multiple linear regression analyses. The dependent variable in these models was the BCSP score, although the independent variables consisted of demographic predictors, including age, gender, region of residence, place of birth, and family national origin. Following the approach used for the first research question, we constructed five separate regression models: one using the total BCSP score as the outcome variable and four additional models analyzing the influence of demographic factors on each of the BCSP section scores (CS history, CS use, CS proficiency, and CS attitudes). This approach allowed us to gain both a comprehensive understanding of how BCSP scores are shaped by social variations and more granular insights into how different aspects of CS experience are influenced by demographic factors. For categorical independent variables, the reference levels were set as follows: female for gender, South and Central American for family origin, Illinois for state of residence, and non-US-born for place of birth. These reference levels serve as baseline categories against which the effects of other groups were estimated.

To ensure the robustness of the results, we checked for key regression assumptions. Specifically, we tested for nonconstant variance (heteroscedasticity) in residuals using the Breusch–Pagan test. If a model exhibited heteroscedasticity, we applied weighted least squares (WLS) regression instead of ordinary least squares (OLS). WLS assigns different weights to observations to stabilize variance, mitigating issues related to heteroscedasticity. After implementing WLS, we rechecked for homoscedasticity to determine whether the issue was resolved. If heteroscedasticity persisted, we applied robust standard errors to account for variance instability and improve the reliability of standard error estimates. In addition to variance checks, we examined the residuals of the models and identified a few data points with large standardized residuals, indicating potential deviations from model assumptions. However, upon further inspection using leverage diagnostics based on a cut-off value of $2(k+1)/n$, where k represents the number of predictors and n is the total number of observations, we found that these data points did not exert undue influence on the model estimates. No data points were identified as influential, confirming that the results were not driven by outliers.

In addition, to check the assumption regarding multicollinearity, we examined the relationships among demographic predictors. Chi-squared tests of independence revealed that state of residence was significantly associated with two other social variables: place of birth ($\chi^2(4) = 16.51, p = .002$) and family national origin ($\chi^2(8) = 73.28, p < .001$). These associations indicated a high degree of overlap in how participants' regional, national, and birthplace backgrounds were distributed. To ensure model stability and improve the interpretability of individual predictors, we decided to exclude state of residence from the regression models. This decision also reduced the risk of overlapping variance obscuring the unique effects of other variables.

Finally, we applied the Bonferroni correction to account for multiple comparisons across both research questions, adjusting the significance threshold to $\alpha = 0.01$ (0.05/5). This correction controls for the family-wise error rate, reducing

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 2. Model comparison using AIC and BIC for random effects selection

Relevant model	Random effects structure	AIC	BIC
Total BCSP	1|Participant	17,867	17,940
	Grammaticality|Participant	17,361	17,447
History BCSP	1|Participant	17,896	17,969
	Grammaticality|Participant	17,387	17,473
Attitudes BCSP	1|Participant	17,869	17,942
	Grammaticality|Participant	17,355	17,441
Proficiency BCSP	1|Participant	17,867	17,939
	Grammaticality|Participant	17,371	17,457
Use BCSP	1|Participant	17,875	17,948
	Grammaticality|Participant	17,361	17,447

Notes: 1|Participant: Random intercepts only; Grammaticality|Participant: Random intercepts + slope for grammaticality.

the likelihood of false positives while ensuring the robustness of our findings. Results were considered statistically significant only if the p-value met this adjusted threshold.

AJT analysis

In this section, we present the results of the mixed-effects models examining how CS experience shapes acceptability judgments. In the model selection process, we found that incorporating a random slope for grammaticality within participants significantly improved model fit compared with the random intercept-only model. The likelihood ratio tests comparing these models yielded highly significant improvements (total BCSP scores: $\chi^2(2) = 510.53$, $p < .001$; history scores: $\chi^2(2) = 512.94$, $p < .001$; attitudes scores: $\chi^2(2) = 518.03$, $p < .001$; proficiency scores: $\chi^2(2) = 499.54$, $p < .001$; use scores: $\chi^2(2) = 518.49$, $p < .001$), indicating that allowing for individual differences in how grammaticality interacts with BCSP scores better captures participant variability. The results also suggest that some participants are more or less influenced by grammaticality. In addition to likelihood ratio tests, we examined the AIC and the BIC values for competing random effects structures across all five BCSP models. For each outcome, the model including random intercepts for participants and items and a random slope for grammaticality within participants yielded lower AIC and BIC values than the random intercept-only model, indicating a consistently better balance between model fit and model complexity (see Table 2). Accordingly, results are reported based on this model specification.

For the first research question, we examined how AJT ratings for Spanish–English CS sentences were influenced by prior CS experience, as measured by BCSP scores. Mixed-effects models were estimated using the total BCSP score and each of

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 3. Model summaries for AJT ratings by BCSP total scores, label status, and grammaticality

Fixed effects	β	*SE*	99% CI	*t*	*p*
Intercept	4.37	.12	[4.05, 4.69]	33.57	<.001***
BCSP total scores	.04	.01	[.02, .06]	4.51	<.001***
Label (with)	.03	.08	[−.17, .23]	0.39	.70
Grammar (gr)	.4	.11	[.13, .67]	3.23	<.001***
BCSP × Label (with)	<.001	.002	[−.01, .01]	−0.16	.87
BCSP × Grammar (gr)	.01	.005	[−.01, .02]	1.62	.11
Label (with) × Grammar (gr)	−.03	.02	[−.43, .36]	−0.23	.82
BCSP × Label (with) × Grammar (gr)	−.005	.004	[−.02, .01]	−1.14	.25
Random effects	**Variance**		***S.D.***		***Corr***
Participant (Intercept)	1.9		1.38		
Participant (Grammar slope)	.7		.84		−.37
Items (Intercept)	.05		.23		
Residuals	1.18		1.08		
Model fit	**Marginal R^2**				**Conditional R^2**
	.09				.68

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

the four section scores (history, use, proficiency, and attitudes) as predictors. The results are summarized in Tables 1–7. Across all models, AJT ratings were significantly higher for grammatical than for ungrammatical sentences, reflecting a consistent and robust main effect of grammaticality.

BCSP total scores showed a significant positive association with AJT ratings ($\beta = 0.04$, $SE = 0.01$, $t = 4.51$, $p < .001$), indicating that participants with greater overall CS experience tended to judge CS sentences as more acceptable. While examining the four BCSP dimensions separately, a clearer pattern emerged. BCSP attitudes scores were strongly and significantly associated with higher AJT ratings ($\beta = 0.10$, $SE = 0.02$, $t = 5.31$, $p < .001$), as were BCSP use scores ($\beta = 0.07$, $SE = 0.01$, t = 5.13, $p < .001$), both remaining statistically significant under the Bonferroni-adjusted threshold of $\alpha = 0.01$. BCSP proficiency scores also exhibited a significant positive relationship with AJT ratings ($\beta = 0.10$, $SE = 0.03$, $t = 2.90$, $p = .004$), meeting the corrected significance threshold. In contrast, BCSP history scores did not significantly predict AJT ratings ($\beta = -0.01$, $SE = 0.01$, $t = -1.68$, $p = .10$), suggesting that mere exposure to past CS events was not a reliable predictor of sentence acceptability.

The interactions between BCSP scores and grammaticality did not reach the Bonferroni-adjusted significance threshold, with the exception of a marginal interaction between CS proficiency and grammaticality ($\beta = 0.04$, $SE = 0.02$,

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 4. Model summaries for AJT ratings by BCSP history scores, label status, and grammaticality

Fixed effects	*β*	*SE*	99% CI	*t*	*p*
Intercept	4.37	.13	[4.03, 4.71]	33.4	<.001***
BCSP history scores	−.01	.01	[−.02, .004]	−1.68	.10
Label (with)	.03	.08	[−.17, .23]	.39	.70
Grammar (gr)	.40	.11	[.13, .68]	3.82	<.001***
BCSP × Label (with)	<.001	.001	[−.002, .004]	.34	.73
BCSP × Grammar (gr)	.003	.003	[−.005, .01]	.81	.41
Label (with) × Grammar (gr)	−.03	.15	[−.43, .36]	−.23	.82
BCSP × Label (with) × Grammar (gr)	−.001	.002	[−.007, .005]	−.54	.59
Random effects	**Variance**		**S.D.**		***Corr***
Participant (Intercept)	2.14		1.46		
Participant (Grammar slope)	.71		.84		−.28
Items (Intercept)	.05		.23		
Residuals	1.18		1.08		
Model fit	**Marginal R^2**				**Conditional R^2**
	.02				.68

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

$t = 2.13$, $p = .04$), which did not meet the corrected threshold for significance. These results indicate that the positive relationship between CS experience and AJT ratings held consistently across both grammatical and ungrammatical sentences. To further assess whether retaining the theoretically motivated interaction terms affected the interpretation of the findings, we compared each full interaction model with a corresponding main-effects-only model. This sensitivity analysis showed that the substantive conclusions were unchanged across model specifications. We therefore report the full interaction models as the primary analyses because they correspond directly to the research questions and experimental design.

BCSP analysis

For the second research question, we examined how BCSP scores are conditioned by demographic factors, including age, gender, place of birth, and family national origin, using multiple regression models. In this article, we present results for the BCSP use and history scores (see Tables 8 and 9), where significant effects were observed; full model results for all BCSP measures are available on the OSF page (https://osf.io/6gcxp/). Overall, the findings indicate that most demographic factors did not show a significant relationship with BCSP scores, suggesting that variation in CS experiences, as measured by the BCSP, is not strongly shaped by

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 5. Model summaries for AJT ratings by BCSP attitudes scores, label status, and grammaticality

Fixed effects	β	*SE*	99% CI	*t*	*p*
Intercept	4.37	.12	[4.06, 4.68]	36.06	<.001***
BCSP attitudes scores	.10	.02	[.05, .15]	5.31	<.001***
Label (with)	.03	.08	[−.17, .23]	.39	.70
Grammar (gr)	.40	.11	[.13, .68]	3.83	<.001***
BCSP × Label (with)	-.01	.005	[−.02, .01]	−1.52	.13
BCSP × Grammar (gr)	.01	.01	[−.02, .05]	.98	.33
Label (with) × Grammar (gr)	−.03	.15	[−.43, .36]	−.23	.82
BCSP × Label (with) × Grammar (gr)	−.01	.01	[−.04, .02]	−1.02	.31
Random effects	**Variance**		**S.D.**		**Cor**
Participant (Intercept)	1.81		1.35		
Participant (Grammar slope)	.71		.84		−.36
Items (Intercept)	.05		.23		
Residuals	1.18		1.08		
Model fit	**Marginal R^2**				**Conditional R^2**
	.12				.68

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

these social variables. Across most models, the coefficients for gender were generally small and statistically nonsignificant, indicating no consistent patterns of association between participants' gender and their BCSP scores. However, age emerged as a significant predictor of BCSP history scores ($\beta = 0.60$, $SE = 0.15$, $t = 4.07, p < .001$), unsurprisingly suggesting that older individuals reported greater past exposure to CS.

Among the few significant findings, place of birth also showed a significant negative relationship with BCSP use scores ($\beta = -3.51$, $SE = 1.21$, $t = -2.89$, $p = .005$), suggesting that individuals born in the USA reported lower levels of CS use compared with those born elsewhere. Additionally, in the model predicting BCSP history scores, place of birth was again a significant predictor ($\beta = 14.31$, $SE = 4.17$, $t = 3.44$, $p < .001$), indicating that US-born participants reported higher past exposure to CS events than non-US-born participants. These findings suggest that although US-born individuals may have had more exposure to CS environments, they engage in CS less frequently than their non-US-born peers.

For all other predictors, including family national origin and regional residence, no consistent effects were observed across the models, suggesting that these factors do not systematically influence BCSP scores. The R^2 values were relatively low across models, indicating that demographic factors alone account for only a small portion of the variability in BCSP scores.

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 6. Model summaries for AJT ratings by BCSP proficiency scores, label status, and grammaticality

Fixed effects	β	*SE*	99% CI	*t*	*p*
Intercept	4.37	.13	[4.04, 4.70]	34.03	<.001***
BCSP proficiency scores	.10	.03	[.01, .18]	2.90	.004**
Label (with)	.03	.08	[−.17, .23]	.39	.70
Grammar (gr)	.40	.10	[.13, .67]	3.85	<.001***
BCSP × Label (with)	.007	.01	[−.01, .03]	.84	.40
BCSP × Grammar (gr)	.04	.02	[−.009, .1]	2.13	.04
Label (with) × Grammar (gr)	−.03	.15	[−.43, .36]	−.23	.82
BCSP × Label (with) × Grammar (gr)	−.02	.02	[−.06, .02]	−1.02	.31
Random effects	**Variance**		**S.D.**		**Cor**
Participant (Intercept)	2.06		1.43		
Participant (Grammar slope)	0.69		0.83		−.35
Items (Intercept)	0.05		0.23		
Residuals	1.18		1.08		
Model fit	**Marginal R^2**				**Conditional R^2**
	.05				.68

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

Discussion

The first research question asks: How are AJT results of Spanish–English CS sentences conditioned by previous experiences with CS, as understood through the BCSP scores? Our results show a positive correlation between the total BCSP scores, encompassing all four sections of the survey, and AJT ratings for grammatical and ungrammatical CS sentences. This suggests that participants with greater experience and engagement with CS are more likely to assign higher acceptability ratings to CS constructions, regardless of grammaticality. We note here that participants do not rate ungrammatical sentences so highly as to conclude that they consider them grammatical. The participants are making a distinction between grammatical and ungrammatical sentences, aligning with other studies (e.g., Badiola et al., 2018). The results support the use of the BCSP and findings from Olson (2025) given the evidence of its effectiveness in capturing participants' real-world experience with CS in a way that aligns with their linguistic behavior, particularly in terms of judgments. To answer our first research question in more detail, we then performed statistical analyses on the four sections that comprise the BCSP.

Not all sections of the BCSP had the same results. The *attitudes* section trended similarly to the overall BCSP score; participants with more positive attitudes were more likely to rate CS sentences higher, which is in line with previous research (e.g., Badiola et al., 2018), further strengthening the idea that attitudes toward CS are

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 7. Model summaries for AJT ratings by BCSP use scores, label status, and grammaticality

Fixed effects	β	*SE*	99% CI	*t*	*p*
Intercept	4.37	.12	[4.06, 4.68]	35.88	<.001***
BCSP use scores	.07	.01	[.04, .11]	5.13	<.001***
Label (with)	.03	.08	[−.17, .23]	.39	.70
Grammar (gr)	.40	.11	[.13, .68]	3.82	<.001***
BCSP × Label (with)	.001	.004	[−.01, .01]	.36	.72
BCSP × Grammar (gr)	.007	.009	[−.02, .03]	.76	.45
Label (with) × Grammar (gr)	−.03	.15	[−.43, .36]	−.23	.82
BCSP × Label (with) × Grammar (gr)	−.003	.007	[−.02, .01]	−.54	.59
Random effects	**Variance**		**S.D.**		**Cor**
Participant (Intercept)	1.83		1.35		
Participant (Grammar slope)	.71		.85		−.34
Items (Intercept)	.05		.23		
Residuals	1.18		1.08		
Model fit	**Marginal R^2**				**Conditional R^2**
	.11				.68

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

Table 8. Model summaries for BCSP use scores conditioned by social variables (WLS estimation)

	Coefficient	*SE*	99% CI	*T*	*p*
Intercept	30.66	2.32	[24.58, 36.73]	13.2	<.001***
Age	−0.08	0.05	[−0.22, 0.05]	−1.63	.10
Family origin (Caribbean)	0.78	2.04	[−4.55, 6.10]	0.38	.70
Family origin (Mexican)	2.9	1.76	[−1.72, 7.51]	1.64	.10
Gender (male)	−2.62	1.42	[−6.32, 1.08]	−1.85	.07
Place of birth (US-born)	−3.51	1.21	[−6.68, −0.34]	−2.89	.005*
Model fit					**R^2**
					0.13

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

a variable that researchers ought to consider. The other BCSP section that followed a similar trend to the total score is *use*. The results show that individuals who reported more frequent use of CS were more likely to provide higher ratings. These results, we imagine, are not controversial. Frequent use could indicate that the participant is

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Table 9. Model summaries for BCSP history scores conditioned by social variables (OLS Estimation)

	Coefficient	*SE*	99% CI	*t*	*p*
Intercept	38.05	7.35	[18.82, 57.28]	5.18	<.001***
Age	0.60	0.15	[0.21, 0.98]	4.07	<.001***
Family origin (Caribbean)	2.15	5.85	[−13.15, 17.45]	0.37	.71
Family origin (Mexican)	4.82	5.30	[−9.05, 18.68]	0.91	.37
Gender (male)	3.20	4.20	[−7.74, 14.15]	0.77	.45
Place of birth (US-born)	14.31	4.17	[3.42, 25.20]	3.44	<.001***
Model fit			R^2		
			.18		

Note: *$p < .01$, **$p < .005$, ***$p < .001$.

part of a community where CS is prevalent, resulting in higher exposure to it, which may also lead to the participant having more favorable attitudes toward CS. Frequent use may also lead the participant to become more sensitive to whether CS sentences are grammatical or not. This last point leads us to the final section of the BCSP—*proficiency*. The results showed a positive and significant relationship between *proficiency* and AJT ratings. In other words, the higher the CS *proficiency* score, the higher the judgments. It follows that if a participant reports high CS proficiency, then they may be more likely to use CS often, which, as explained above, might lead to higher AJT ratings. However, although the results have shown that *proficiency* is a significant predictor of higher ratings, we believe the critical relationship lies elsewhere: the statistically significant distinction between ungrammatical vs. grammatical CS sentences. High proficiency in CS necessarily means that the participants are sensitive to grammatical constraints, allowing them to distinguish whether a CS sentence is acceptable or not. However, although we can assert a positive relationship, we cannot do the same for the opposite. That is, future studies could investigate whether low CS *proficiency* scores predict that participants will not be able to distinguish between ungrammatical and grammatical CS sentences.

The remaining section, *history*, did not exhibit a significant relationship with acceptability judgments. With *history*, the BCSP's goal is to consider exposure to CS in a long-term context, as Olson's rationale is that CS exposure is not constant throughout a speaker's life. Our results indicate that history is not necessarily a reliable predictor of judgments. This result appears to be counterintuitive; one could assume that a long history of exposure and use of CS would lead to higher acceptability judgments. However, the *history* section of the survey aims to measure long-term and past exposure to CS, including the contexts (e.g., school, home, and work). Although history may be critical in other experimental procedures, the saliency of CS is perhaps more important at the time of testing judgments. Consider the other sections that yielded significant results: *use* and *attitudes*. These sections tap into the present-day engagement of CS. Thus, although *history* reports past engagement with CS, it does not necessarily mean that they are currently exposed to

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

CS, or that they currently have positive attitudes toward the practice. Future research may consider testing the interitem correlations or conducting a cluster analysis *within* each BCSP section to determine whether certain questions are better predictors of AJT ratings.

These findings support not only a usage-based approach to bilingual language processes, but also contemporary approaches that directly address the inherent heterogeneity within bilingual populations. We show that variation in CS use, proficiency, and attitudes significantly influences AJT ratings, and we posit that taking into account the myriad experiences, proficiencies, attitudes, and daily linguistic experiences of US Latinx bilinguals is crucial in achieving a more comprehensive understanding of acceptability judgments. We also note that, although not the primary focus of the current analyses, the methodological manipulation of the added reggaetón labels did not influence AJT results in our sample. This follows Callesano et al. (2025), whereby source labels do not have uniform influence on acceptability judgments.

Let us now move on to the second research question, which asks: How are the results of the BCSP conditioned by age, gender, region of residence, place of birth, and family national origin? Our results indicate that most demographic factors did not show a significant relationship with BCSP scores, suggesting that the BCSP results are not influenced by these variables. Here, we would like to highlight that these results pertain only to the CS structures tested here, and they do not mean that nonsignificant variables are irrelevant. There have been studies that show that family national origin and regional residence, for example, do have an effect in specific CS structures. As mentioned above, Balam et al. (2020) found an asymmetry in the acceptance of bilingual compound verbs between communities in Belize vs. the US Southwest (see Olson [2025] for a corroboration of these results). Future studies should therefore test a variety of CS structures to determine if social variables have an effect on their acceptability. Let us continue to the variables that did have an effect. For example, *age* was a significant predictor of the *history* subsection scores, indicating that older individuals reported greater past exposure to CS. This result intuitively makes sense; for example, an 80-year-old early bilingual likely has a more extensive history with CS than a 20-year-old early bilingual.

Another significant relationship was between *place of birth* and the CS *use* sections. The results show that individuals born in the US reported lower levels of CS use compared with those born abroad. One might expect that those born in the USA, where contact between Spanish and English is the most prevalent, would report greater use of CS. However, this assumption is complicated by the stigmatization of CS in the USA due to monolingual ideologies (the idea that we must keep Spanish pure and separate from English); as a result, bilingual speakers in the USA may be less likely to use CS or may have negative attitudes toward the practice (e.g., Rangel et al., 2015). The higher levels of CS use among those born outside of the USA might also be a reflection of the acculturation processes and greater integration of English into everyday speech. Furthermore, the music genre implemented into our study might also be generally increasing global perceptions of CS given that CS has become a more common practice in reggaetón lyrics in recent years (Monteagudo, 2020). Lastly, CS in other Spanish-speaking countries, such as Mexico, might be seen as a sign of prestige—no longer is English encroaching upon

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Spanish; rather, it is a sign of cosmopolitanism (Baumgardner, 2006). We also note that our results are fairly in line with those reported by the Pew Research Center, which report that 36% of US-born Latinos say they rarely/never use "Spanglish" with friends and family, although 34% of foreign-born Latinos say they rarely/never use "Spanglish." More foreign-born Latinos also report using "Spanglish" sometimes as compared with those born in the USA. However, equal percentages of foreign and US-born Latinos report using "Spanglish" extremely often/often (Pew Research Center, 2023). The significant relationship between *place of birth* and *history* further supports this idea.

Our results show that US-born participants reported higher past exposure to CS than non-US-born participants. This is another expected result because, in the USA, Spanish and English are in constant contact; thus, CS is more prevalent than in a Spanish-speaking country like Mexico, where CS does indeed occur, but not at the same frequency. The previous two results—US-born participants report less use but more exposure to CS—might seem contradictory at first, but let us recall that CS does not occur in a vacuum. Other variables may also affect the use, even if the history remains constant. These could be attitudes, which might play a stronger role in mediating CS behavior more generally than overall exposure to CS, and another confounding factor might be current place of residence, which has shown to affect acceptability of specific structures, such as bilingual compound verbs (Balam et al. 2020). Our data are geographically distributed by US state, and not by major US Latinx city (e.g., Los Angeles vs. Miami), and location of residence can influence CS use, where one might posit that bilinguals in major cities (e.g., Chicago) are more likely to engage in CS than bilinguals in other regions of the same state with smaller Latinx populations (e.g., central Illinois).

Overall, our analyses have further highlighted the importance of considering context in bilingual grammatical acceptability research. By "context" here, we refer to the sociocultural environment associated with language use, reggaetón music in this case, and the individualized CS experiences of the survey respondents—the BCSP scores. In the current analysis, the BCSP scores at various levels played a significant role in conditioning the CSAJT findings, providing further evidence for usage-based and holistic approaches in linguistics research. This makes the BCSP useful not only for capturing bilingual heterogeneity, but also for representing bilinguals' cognitive and linguistic experiences, which shape acceptability judgments. Our work has illustrated, albeit minimally, some variations in self-reported CS experiences using data from a national sample, with place of birth and age playing significant roles. The lack of major variation in CS experiences was an unexpected finding, as our informal survey evaluators between Illinois and Texas provided the authors with some feedback that indicated variables CS experiences. What we argue here is that although CS experiences (as measured by the BCSP) may not vary significantly in our sample, other measures that address *how* different bilingual communities engage in CS might be more insightful. That is, as we heard from our informal survey evaluators, some communities prefer to switch languages in shorter syntactic chunks (akin to congruent lexicalization [Muysken, 1997]), although others prefer to produce longer stretches of discourse in one language before switching to the next. CS experiences at the level of discourse, then, might provide a more nuanced view into the variable CS experiences of different groups

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

and individuals, and could be an important factor in how language use conditions language perception.

Our article is not without its limitations. Future work should expand on the syntactic structures beyond weak vs. strong subjects used here to increase the generalizability of results. Future work may also consider including either switch direction or identifying different artists as the source of the switch, as well as important semantic differences between English and Spanish within CS production. Furthermore, the interpretability of the label-related findings could be improved by a follow-up study in which an additional label is used. Although no label-related finding was noted in here, an effect of the reggaetón label is reported by Callesano et al. (2026), and a possible follow-up study could help to ensure that the label-related findings are influenced specifically by the reggaetón label, as opposed to the presence of any label more generally. In this study, we used CloudReserach to recruit our participants, limiting the responses to those who had access to the internet, computers, and higher levels of literacy. Future studies should include a broader spectrum of participants, and in particular focus on major Latinx cities.

Conclusion

This study provides novel empirical evidence linking the BCSP to AJTs, advancing our understanding of how engagement and exposure to CS can inform acceptability judgments. Our results showed a positive correlation between the overall BCSP score and acceptability judgments; those with higher BCSP scores gave higher ratings. More specifically, however, CS use, attitudes, and proficiency were also significant predictors of higher judgments. These results support cognitive models of bilingualism that incorporate individual variation and heterogeneous bilingual experiences (e.g., López et al., 2023), following usage-based accounts of bilingual grammatical intuitions (e.g. Backus, 2015). Additionally, although demographic variables had limited explanatory power for overall BCSP scores, *place of birth* and *age* showed nuanced effects. Taken together, our results underscore the value of integrating instruments like the BCSP into formal linguistic research, and for continued interdisciplinary approaches that account for both structure and experience in bilingual language use, aligning with trends in usage-based and socio-cognitive models of bilingualism, which argue that linguistic knowledge is shaped not only by internal grammar but also by lived language experience and identity.

Competing interests. The author(s) declare none.

Replication package. The materials, data, code and analyses for the results reported in the paper are available at: https://osf.io/6gcxp/.

Note

[1] With fusions of Jamaican dancehall, reggae *en español*, and hip hop, reggaetón music is characterized by fast paced lyrics, dembow beats, and crossing national boundaries from the Caribbean, Central America, and the USA (Marshall, 2009). The genre became immensely popular on global scales in the early 2000s and has recently been dominated by Bad Bunny, who won a Grammy for Album of the Year (2026) and headlined the 2026 National Football League Superbowl.

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

References

Anderson, T. K., & Toribio, A. J. (2007). Attitudes towards lexical borrowing and intra-sentential code-switching among Spanish-English bilinguals. *Spanish in Context*, **4**(2), 217–240. https://doi.org/10.1075/sic.4.2.05and

Backus, A. (2015). A usage-based approach to code-switching: The need for reconciling structure and function. In G. Stell,& K. Yakpo (Eds.), *Code-switching between structural and sociolinguistic perspectives* (pp. 19–38). De Gruyter. https://doi.org/10.1515/9783110346879.19

Badiola, L., Delgado, R., Sande, A., & Stefanich, S. (2018). Code-switching attitudes and their effects on acceptability judgment tasks. *Linguistic Approaches to Bilingualism*, **8**(1), 5–24. https://doi.org/10.1075/lab.16006.bad

Balam, O., Del Carmen Parafita Couto, M., & Stadthagen-González, H. (2020). Bilingual verbs in three Spanish/English code-switching communities. *International Journal of Bilingualism*, **24**(5–6), 952–967. https://doi.org/10.1177/1367006920911449

Balam, O., & Shelton, J. (2023). The use of Spanglish in Latin rap music. *Bilingual Review/La Revista Bilingüe***35**(1), 32–52.

Baumgardner, R. J. (2006). The appeal of English in Mexican commerce. *World Englishes*,**25**(2), 251–266.

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, **95**, 173–189.

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. Web. 20 Jan.<https://sites.la.utexas.edu/bilingual/>.

Blokzijl, J., Deuchar, M., & Parafita Couto, M. C. (2017). Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages*, **2**(4), 20. https://doi.org/10.3390/languages2040020

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta psychologica*, **193**, 160–170.

Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2020). Cognitive control in bilinguals: Effects of language experience and individual variability. *Bilingualism: Language and Cognition*, **23**(1), 219–230. https://doi.org/10.1017/S1366728918001086

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, **82**(4), 711–733.

Callesano, S., & Carter, P. M. (2019). Latinx perceptions of Spanish in Miami: Dialect variation,personality attributes and language use. *Language & Communication*, **67**, 84–98.https://doi.org/10.1016/j.langcom.2019.03.003

Callesano, S., Delgado, R., Ma, W., & Trybulec, E. (2026). Social variation of code-switching acceptability. *International Journal of Bilingualism*, **0**(0). https://doi.org/10.1177/13670069251404838.

Callesano, S., Delgado, R., Ma, W., Trybulec, E., & Reyes, G. (2025). Contextualizing code-switching: Effects of reggaetón labels and listening frequency in an acceptability judgment task. *Frontiers in Communication*, **9**. https://doi.org/10.3389/fcomm.2024.1490299

Carter, P. M., & Callesano, S. (2018). The social meaning of Spanish in Miami: Dialect perceptions and implications for socioeconomic class, income, and employment. *Latino Studies*, **16**(1), 65–90. https://doi.org/10.1057/s41276-017-0105-8

Dewaele, J. M., & Li, W. (2014). Intra- and inter-individual variation in self-reported code-switching patterns of adult multilinguals. *International Journal of Multilingualism*, **11**(2), 225–246. https://doi.org/10.1080/14790718.2013.878347

González-Vilbazo, K., Bartlett, L., Downey, S., Ebert, S., Heil, J., Hoot, B., Koronkiewicz, B.& Ramos, S. (2013). Methodological considerations in code-switching research. *Studies in Hispanic and Lusophone Linguistics*, **6**(1), 119–138. https://doi.org/10.1515/shll-2013-1143

González-Vilbazo, K., & Koronkiewicz, B. (2016). Tú y yo can codeswitch, nosotros cannot: Pronouns in Spanish-English codeswitching. In R. E. Guzzardo Tamargo, C. Mazak,& M. Parafita Couto (Eds.), *Spanish-English codeswitching in the Caribbean and the US* (pp. 237–260). John Benjamins.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Hay, J.,& Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, **48**(4), 865–892. https://doi.org/10.1515/ling.2010.027

Hayes, E. N. (2022). The sociophonetics and lyrical codemixing of hot Latin songs: A comparison of the spontaneous speech and the artistic performance speech of the top two artists in the global market, bad bunny and J Balvin [Doctoral dissertation, The University of Alabama].

Herring, J. R., Deuchar, M., Couto, M. C. P., & Quintanilla, M. M. (2010). 'I saw the madre': Evaluating predictions about codeswitched determiner-noun sequences using Spanish–English and Welsh–English data. *International Journal of Bilingual Education and Bilingualism*, **13**(5), 553–573.

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2019). Predicting executive functions in bilinguals using ecologically valid measures of code-switching behavior. In I. A. Sekerina, L. Spradlin,& V. Valian (Eds.), *Bilingualism, executive function, and beyond: Questions and insights* (pp. 181–205). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.57.12hof

Koronkiewicz, B. (2014). Pronoun categorization: Evidence from Spanish/English code-switching. [Dissertation/doctoral thesis]. Chicago, IL: University of Illinois at Chicago.

López, B. G., Luque, A., & Piña-Watson, B. (2023). Context, intersectionality, and resilience: Moving toward a more holistic study of bilingualism in cognitive science. *Cultural Diversity & Ethnic Minority Psychology*, **29**(1), 24.

López, B. G., & Vaid, J. (2018). Fácil or A piece of cake: Does variability in bilingual language brokering experience affect idiom comprehension?*Bilingualism: Language and Cognition*, **21**(2), 340–354. https://doi.org/10.1017/S1366728917000086

López, B. G., Lezama, E., & Heredia, D. J. (2019). Language brokering experience affects feelings toward bilingualism, language knowledge, use, and practices: A qualitative approach. *Hispanic Journal of Behavioral Sciences,* **41**(4).

Macswan, J. (2005). Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism."*Bilingualism: Language and Cognition*, **8**(1), 1–22. https://doi.org/10.1017/S1366728904002068

Marshall W. (2009). From música negra to reggaeton latino: the cultural politics of nation, migration, and commercialization. In R. Z. Rivera, W. Marshall, & D. P. Hernandez(Eds.), *Reggaeton* (pp. 19–76). Duke University Press.

Monteagudo, M. (2020). Spanglish code-switching in Latin Pop music: functions of English and audience reception [Master's thesis, University of Oslo].

Muysken, P. (1997). Code-switching processes: Alternation, insertion, congruent lexicalization. In M. Pütz (Ed.), *Language choices: Conditions, constraints, and consequences* (pp. 361–380). John Benjamins Publishing Company.

Myers-Scotton, C. (1993). Duelling languages: Grammatical structure in codeswitching. Clarendon Press. https://doi.org/10.1093/oso/9780198240594.001.0001

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, **18**(1), 62–85. https://doi.org/10.1177/0261927X99018001005 (Original work published 1999)

Olson, D. J. (2022). The Bilingual Code-Switching Profile (BCSP). *Linguistic Approaches to Bilingualism*, **14**(3), 400–433. https://doi.org/10.1075/lab.21039.ols

Olson, D. J. (2023). A systematic review of proficiency assessment methods in bilingualism research. *International Journal of Bilingualism*, **28**(2), 163–187. https://doi.org/10.1177/13670069231153720

Olson, D. J. (2025). Bilingual language experience and code-switching acceptability judgments:A constructive replication of the work by Stadthagen-González et al. (2019), Balam et al. (2020), and Stadthagen-González et al. (2018). *International Journal of Bilingualism*, **29**(6), 1724–1754.

Parafita-Couto, M. C., Romeli, M. G., & Bellamy, K. (2021). Code-switching at the interface between language, culture, and cognition. Lapurdum. Available online: https://halshs.archives-ouvertes.fr/halshs-03280922/

Pew Research Center. (2023, September 20). Latinos' views of and experiences with the Spanish language. https://www.pewresearch.org/race-and-ethnicity/2023/09/20/latinos-views-of-and-experiences-with-the-spanish-language/

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics*, **18**(7–8), 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Rangel, N., Loureiro-Rodríguez, V., & Moyna, M. I. (2015). "Is that what I sound like when I speak?" *Spanish in Context*, **12**(2), 177–198. https://doi.org/10.1075/sic.12.2.01ran

Regan, B., & Toribio, A. J. (2026). '*Como está hablando con su mamá, está bien que haga el code-switching*': Social perceptions of situationally-based code-switching in West Texas. *Journal of Multilingual and Multicultural Development*, 1–27. https://doi.org/10.1080/01434632.2026.2612963

Rojas-Berscia, L. M. (2024). Dynamic linguistics: An aperitif. *Nota Bene*,**1**, 88–114. https://doi.org/10.1075/nb.00006.roj

Sprouse, J. (2018). Acceptability judgments and grammaticality, prospects and challenges. In N. Hornstein, H. Lasnik, P. Patel-Grosz,& C. Yang (Eds.), *Syntactic structures after 60 Years: The impact of the Chomskyan Revolution in Linguistics* (pp. 195–224). De Gruyter Mouton. https://doi.org/10.1515/9781501506925-199

Sprouse, J.,& Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: AJournal of General Linguistics*,**2**(1), 14. https://doi.org/10.5334/gjgl.236

Yim, O., & Bialystok, E. (2012). Degree of conversational code-switching enhances verbal task switching in Cantonese–English bilinguals. *Bilingualism: Language and Cognition*, **15**(4), 873–883. https://doi.org/10.1017/S1366728912000478

Zentella, A. C. (1997). Latino youth at home, in their communities, and in school: The language link. *Education and Urban Society*, **30**(1), 122–130. https://doi.org/10.1177/0013124597030001008 (Original work published 1997).

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.

Appendix



Figure 1. Screenshot of Qualtrics survey. Example of grammatical CS without reggaetón source label.

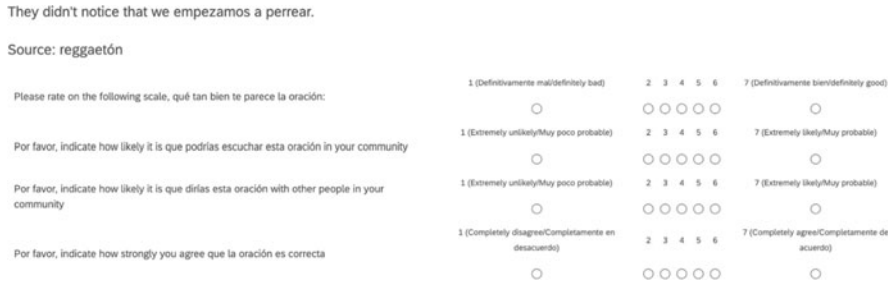


Figure 2. Screenshot of Qualtrics survey. Example of ungrammatical CS with reggaetón source label.

Cite this article: Delgado, R., Callesano, S., & Ma, W. (2026). Variable bilingual experiences and their influence on a contextualized acceptability judgment task. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716426100873

Downloaded from https://www.cambridge.org/core. 09 Nov 2026 at 23:25:30, subject to the Cambridge Core terms of use.