CC: editorial@labjournal.org

LAB-23057 (Linguistic Approaches to Bilingualism)
"Home and School Exposure and Age Effects in the Heritage Language Acquisition of the Spanish Volitional Subjunctive"
by Patrick D. Thane

Dear Patrick Thane,

We write today with reference to your submission to LAB entitled "Home and School Exposure and Age Effects in the Heritage Language Acquisition of the Spanish Volitional Subjunctive".

Based on the recommendation of the reviewers (included herein, or, alternatively, available by logging in to the editorial manager, clicking on "action links", then on "view attachments") as well as our own reading of the manuscript, we would like to invite you to revise and resubmit the article. We find the topic extremely interesting, but we agree with the reviewers that there are some issues that we feel need to be addressed in a revision.

In particular, please note the reviewers' comments regarding the following:

> Incorporation of a dedicated analysis section that motivates the analyses used and includes details about model fitting procedures;

> Reviewer 1's comments regarding the link between the data and Putnam and Sánchez' (2013) claims. Specifically, their request for more in-depth consideration of features in terms of how they are conceptualized and how the data align with more than one possible outcome.

Together with your revised manuscript, please send us a letter copying all the reviewers' comments, followed by an explanation of exactly how you have addressed each comment and suggestion. Please make sure that your manuscript follows the LAB stylesheet (https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.benjamins.com%2Fseries%2Flab%2Flab_style_sheet.pdf&data=05%7C02%7Cpthane%40umass.edu%7Cc660338473064fe014d608dc2b1bcc9c%7C7bd08b0b33954dc194bbd0b2e56a497f%7C0%7C0%7C638432646114217774%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=7rMZNMAitOHUHp3wV6RRHHtRu4xd3qJheu8K8vfp%2B2g%3D&reserved=0). Manuscripts which do not adhere to these guidelines will not be accepted for publication.

Please note that an invitation to revise and resubmit does not entail that the next version, or any subsequent version, will be accepted for publication. The revised version will be evaluated anew, and it is sometimes the case that manuscripts which are still judged not to be acceptable after revisions have been made are rejected. Additionally, the clarifications and/or changes in the revision may uncover new or additional issues that lead to rejection of the manuscript.

To submit your revision, please go to https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.editorialmanager.com%2Flab%2F&data=05%7C02%7Cpthane%40umass.edu%7Cc660338473064fe014d608dc2b1bcc9c%7C7bd08b0b33954dc194bbd0b2e56a497f%7C0%7C0%7C638432646114217774%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=71NR5Pj%2BGjEf4g17Wi2A8N8U8O1thavH%2FaSASh2rvP8%3D&reserved=0 and log in as an author. There, you can access the menu

QTbuvTT%2FaSAGnZryT-O%3D&reserved=0 and log in as an author. There, you can access the menu item 'Submission Needing Revision'.

LAB requires you to make your materials and/or data freely available to enhance the visibility and reproducibility of your study. If you do this, your article will be awarded recognition from the Center for Open Science via an 'Open Materials' badge and/or an 'Open Data' badge which will be clearly displayed next to your article. This is a public kitemark, recognizing that you are practising open science. To qualify for a badge, your materials (and/or data) must be on an open repository that is recognised by the Center for Open Science, such as IRIS (https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.iris-database.org%2F&data=05%7C02%7Cpthane%40umass.edu%7Cc660338473064fe014d608dc2b1bcc9c%7C7bd08b0b33954dc194bbd0b2e56a497f%7C0%7C0%7C638432646114217774%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C0%7C%7C%7C&sdata=lhDK1rtiyw%2BTfbHiLJYoK%2B28CGM7efaGaRfIQozLAhY%3D&reserved=0).

Thank you once again for your submission, and we are looking forward to reading your revised manuscript as soon as possible.

Best wishes,

The Editors

Linguistic Approaches to Bilingualism

Comments:

Reviewer #1: This manuscript documents the results of a study targeting the acquisition of the Spanish subjunctive mood by 57 English-dominant heritage speakers in fifth, seventh, and eight grades (e.g., ages 10-14), with some of these participants being enrolled in a dual-language immersion (DLI) program. These results were compared with those a comparison/baseline group of Spanish-dominant adults (n=18). Unsurprisingly, the adults used subjective mood categorically, while the groups of HL-speaking children were must more variable in their production and comprehension. Ultimately, the authors (hereafter A) suggest that the use and asymmetric(al) performance between the tasks among these different groups of speakers largely supports Putnam & Sánchez (2013; P&S hereafter) approach to HL-acquisition. This manuscript is well-written and generally well-structured, making a clear case for the research questions raised and highlighting the relevance of this individual study in the context of ongoing research on the acquisition/development of HLs more generally. The experimental tasks associated with this manuscript are incredibly well-designed in my view, and I have little criticism or additional commentary to offer in this area. My primary concerns with this manuscript are directed towards the discussion of how these findings relate to P&S's claims. These points notwithstanding, I don't feel that A will encounter any significant challenges or difficulties addressing these questions and concerns, and, as a result, I feel that a verdict of "Accept pending Major Revisions" suffices in this case.

Questions & concerns:

1. Please be sure to cite the Putnam & Sánchez (2013) piece in your references. It's quite surprising that it's not there...
2. The treatment/discussion of "features" in fn. 1 requires further elaboration, even if the primary focus of this manuscript isn't directly involved with the advancement of any particular theoretical claims. One reference that is worth consulting in this respect is Lohndal & Putnam (2021; hereafter, L&P), who provides a more detailed treatment of the conceptualization of "features" that is largely compatible with P&S's (2013) proposals. As such, the notion of "feature" seems to function as the fundamental unit of linguistic structure in this paper, hence, perhaps it should not appear in a footnote, but rather in the main body of prose.
3. Page 10, lines 22-28: Two points are in order here: First, building upon my previous comment, a more detailed treatment of "features" in this paper would allow A to say something that the the features

detailed treatment of "features" in this paper would allow A to say something that the the features responsible for subjunctive mood (in connection with the morphology that 'realizes' these features). Second, it sounds a bit awkward to say something along the lines of a grammatical feature (or set of features) being "less like to reassemble"; rather, I think that something along the lines of "would be less likely to be impacted", since ultimately what's going on here is a resistance to reassemble (or, perhaps, different associations with allomorphy).

4. A question concerning the interpretation of these experimental results involves whether or not features have been "reassembled", or perhaps, their associate with allomorphy that expresses subjunctive mood has become more variable. Although the P&S (2013) model would support both, an important contribution that A could make here, along the lines of what L&P propose on p. 11 (6) of their article, is that the "loss" or "reconfiguration" of features is one of several outcomes. What some of these groups/individuals in this study may be doing is exhibiting variable feature-morphology associations. Again, although the larger focus and purpose of this paper is not a theoretical contribution per se, it's worth mentioning this, especially since this is in line with the general position offered by A in their interpretation of their own results.

5. Page 25, lines 4-24: The question raised by A re: whether or not the P&S (2013) model is adaptable to children, or at the very least requires some deeper thought. I would suggest also taking a look at Putnam et al. (2018) for an expansion of this topic. Surely, this is likely something that will receive short shrift in this piece, but it's worth acknowledging how this could contribute to a larger a discussion on these and related matters.

References

Lohndal, T., & Putnam, M.T. 2021. The tale of two lexicons: Decomposing complexity across a distributed lexicon. Heritage Language Journal, 18(2), 1-29.

Putnam, M.T., Kupisch, T., & Pascual y Cabo, D. (2018). Different situations, different outcomes: Heritage grammars across the lifespan. In: D. Miller, F. Bayram, J. Rothman, & L. Serratrice (eds.), Bilingual cognition and language: The state of the science across its subfields (pp. 251-281). Amsterdam: John Benjamins.

Reviewer #2: LAB review

General comments:
This study is ambitious in scope because it considers age and type of education along with frequency of use. I can't recall a subjunctive study that has considered all of these together, so this is appreciated. That said, a sample size of 57 students is small, especially because there are no more than 19 students in each cell. The authors do acknowledge the sample size as a limitation, but only in one sentence in reference to the lack of a schooling effect. I agree with what they say, but I would argue more generally that we need follow-up studies with larger samples to confirm all of the findings of the present study.

I recommend including an analysis section in between the methods and results. I am not completely clear on the dependent variables (see specific comment below for p. 18). And if the models are using multiple observations per person (as opposed to a tallied score for one or both tasks), I wonder if would be more appropriate to nest observations (e.g., responses on a task) within individuals. Given the relatively small sample size, an analyses section could also explain the suitability of the statistical models for this number of participants and describe any preliminary tests that may have been conducted to determine the suitability of these models.

Lit review
From the abstract, lit review and the research questions, it becomes clear that the study focused on the volitional subjunctive in nominal clauses, which is perfectly reasonable. However, I think that it would help readers—especially those who do not have a background in Spanish—to explain *why* that subjunctive. This doesn't have to be extended, just enough to clarify that Spanish does use the present subjunctive elsewhere (including adjectival and adverbial clauses), but that volitional contexts in nominal

subjunctive elsewhere (including adjectival and adverbial clauses), but that volitional contexts in nominal clauses are where we do not have to worry about regional variations or a speaker's personal belief about what is being expressed. To this end, footnote 4 could be incorporated into the text of the lit review. You could also bring in Blake (1983), in the sense that children start to acquire the volitional subjunctive early, so this use (at least in theory) would therefore be less vulnerable to attrition.

Methods
1. I understand the choice to hold morphological regularity constant. However, it is important to acknowledge that verbs that have a stem change in the subjunctive are likely to be more salient to students in both oral and written language. For example, the 3s present subjunctive form tenga (indicative counterpart tiene) may stand out more than the 3s present subjunctive form hable (indicative counterpart habla). See Collentine's work, in particular his 1997 piece in Spanish Applied Linguistics. Collentine's participants are traditional FL learners, but the difference in salience is relevant to heritage learners as well.

Why do I bring this up? Because some very common verbs have changes in the stem (e.g., forms of tener, poner, salir). One can argue that by holding morphological regularity constant, you are hindering your participants; that is, by not including some of the subjunctive forms that they are most likely to hear and some of the forms that are most likely to stand out to them, you are reducing their chances for success. Consequently, although this choice was beneficial in that it eliminated a variable from consideration, the limitations of this choice also need to be acknowledged.

2. Regarding the MLE school, many English-language middle schools in the US offer Spanish as an elective (typically designed for traditional FL students). Did the school that the MLE 7th and 8th graders attended offer such courses? If so, were any of the participants enrolled in those courses? This is important to note: even though middle school Spanish classes don't typically teach the subjunctive explicitly, these classes would be another source of input for any MLE students. I understand that the frequency of language use measure asks about school, but it would be helpful to include this information so readers have an idea of whether Spanish at an MLE school may include some classroom instruction.

Along the same lines, I noticed on p. 24 that it says that the DLI group had their input decrease to just 1 class a day. Was this class a Spanish as a foreign/world language class or was it more a of a language arts (reading and composition) class in Spanish? Were they attending schools designated as DLI where the proportion of Spanish instruction shifted, or were these children who attended a DLI elementary but moved into an MLE middle school? These details have important implications for Spanish use at school, so I would like to see more information on this.

Specific comments:
(page numbers are the ones in the upper right corner, put in by the author, not the page of the pdf)

p. 4, first full paragraph:
a) Did Gathercole (2002) really use experimental methods? Unless she randomly assigned students to conditions (which seems unlikely), she didn't. Maybe you mean quasi-experimental (e.g., comparison of 2 groups of people that are not randomly assigned to conditions). Maybe you just mean longitudinal and/or comparative.
b) I tried to look up the specific Gathercole article to verify the methodology used, but the reference is not in the bibliography.
c) Also, I'm not sure why it's ironic that only Gathercole has done this. Is it that we have a gap in the research that only she has sought to fill? This should be clarified.

p. 9: SDBA= Spanish-Dominant Bilingual Adults, correct? Please spell out the abbreviation upon its first use.

p. 12: MLE-7/8 n = 11, as shown in the table. It says n = 25 in the text, but the table shows that it is the total for MLE at all grade levels.

p. 13: "The children enrolled in the monolingual school were matched for age, socioeconomic status, and family background with those in DLI." How so? This doesn't look like a study with matched pairs because

family background with those in DLI." How so? This doesn't look like a study with matched pairs because Table 1 show us different numbers of participants in each cell. Was it one-to-many matching? Reading the rest of the paragraph, my guess is that the research was conducted at MLE and DLI schools that had similar demographic profiles (i.e., similar socioeconomic status (SES) and family backgrounds), though I'm still unclear on the age part beyond the fact that recruiting from the same grade (e.g., 5th at MLE and DLI schools) will yield participants with similar ages. Choosing schools with similar profiles—while commendable—is not "matching" as it is understood in statistical analyses. I'd like to see a clearer explanation of this matching.

p. 13: "While only some children's parents spoke English, all spoke Spanish, and all participants were predominantly exposed to Spanish at home (see Table 2)." With regard to "some children's" and "all", are we talking about the children at the school generally, or the participating children? It seems like the latter, but this should be clarified.

p. 13: "The SDBAs… represent a source of input for the HS groups." How so? Where they recruited from the same community where the schools were located? Are some of them parents/family members of child participants? Do some of them teach at the DLI school? Please elaborate.

p. 13: Please explain how frequency of use of Spanish is being measured. I think that you get to this later, so it may be a matter of moving text up or telling the reader that these measures are described later on.

p. 14: Please explain the proficiency measure used in Table 2. It is not the full DELE, given that it only has a maximum of 18 points. Is this a subset of the DELE? If so, what types of items were included? (It seems like a lot of DELE points for volitional subjunctive only. Is it for all subjunctives or some other combination of items?) Reading on further, it looks like it was the BESA. Again, this is a matter of moving text up or telling the reader that these measures are described later on.

Also, if Table 2 is reporting the BESA results and 4 items were not reported, shouldn't the maximum score on the table be 14?

p. 16: I'm a bit uneasy about the items with creer being used to test indicative mood, just because we can see some mood variation in negated epistemics like no creer que, which could carry over into the affirmative counterparts if the speaker wants to emphasize doubt. A safer bet would have been testing after tener que, saber que, or ver que. I suggest acknowledging this in the limitations.

p. 17: "In the HSs' production data, the children produced the subjunctive in a total of 131/426 instances (30.7%), and alternative forms in the remaining 295 instances (69.2%)." If this is the case, why does every child group in Figure 1 have production percentages over 30.7%? I understand that the 30.7% is for all children and Figure 1 shows the subsets, but I don't see how every group doing better than 30.7% can mathematically also work out to 30.7% for all children.

p. 18: "In both models, the suppliance of the expected mood inflection was incorporated as the dependent variable…" I'd like to see the phrase clarified. For any participant, one can calculate a preference (receptive) score and a productive score. So, is the DV a combined receptive + productive score? Or is this more of an odds ratio, that is, the odds that participants will answer correctly? This is one thing that could be included in an analysis section.

p. 21: "some who showed great variability on the proficiency test" Do you mean that these 6 students showed a wide range of scores on the proficiency test?

Footnote 9: Do you mean to refer the reader to footnote 8? Both notes may work better as notes on tables rather than footnotes to the text.

p. 26: Like my comment on Gathercole (2002) above, I wonder if "experiments" is the most accurate term here.

p. 26: Although every verb in the present study was a regular -er verb, the authors could look at

p. 26: Although every verb in the present study was a regular -ar verb, the authors could look at frequency of those verbs. That is, did students perform better on the items featuring more frequent verbs?

p. 26: Regarding the discrepancy between the findings of the present study and those of Potowski (2007a), how many of the subjunctive items in the latter had stem changes? This could be another factor in addition to the amount of Spanish used for instruction at the DLI school.

p. 28: I agree that we need to consider the importance of output in Spanish (which a DLI setting does not necessarily guarantee) and the role of input quality when considering why we didn't see differences between MLE and DLI students. Something else to consider is the nature of language use questionnaires. Language use at home can vary depending on the interlocutor(s) and the topic of conversation, and it's difficult for a brief questionnaire to capture all of that complexity.

Minor errors
p. 2, line 42: HL acquisition are > HL acquisition is
p. 3, line 11: empiricalize > operationalize
p. 10, line 18: regards to > regard to
p. 10, line 23: exposure to and processing > exposure and processing (or maybe there are some missing words?)
I see other minor errors and recommend a thorough proofreading of the manuscript.

Reviewer #3: Review of the manuscript entitled "Home and school exposure and age effects in the heritage language acquisition of Spanish volitional subjunctive", submitted to the journal Linguistic Approaches to Bilingualism.

This paper presents a topical study in a still under-researched area of heritage language acquisition. The subjects are school-aged English dominant heritage speakers of German and the author(s) investigate receptive and productive knowledge of the Spanish volitional subjunctive. They additionally control for language exposure in school (bilingual immersion versus English program) and at home as well as age (school year 5 versus school years 7/8). Although this presents an interesting contribution, there are a number of points that I find worth pointing out. I invite the author(s) to consider these when revising the paper. A few points are listed below; the remaining points are directly are added as comments into the manuscript.

-     Please carefully proofread the paper again; a number of (grammatical) errors and errors in Figures have been pointed out in the PDF. In addition, please introduce all abbreviations at first mention.
-     Some sentences are very long, run over multiple lines, and are separated by numerous citations within - this makes reading and understanding somewhat cumbersome. Perhaps consider rephrasing some of these to produce as smoother writing style.
-     I think some of the contextualization needs to be a bit more in depth. For example, who are your SDBAs? And please provide more context to the type of schooling - this information was bit scattered here and there and appeared a bit inconsistent. Please check and provide a more coherent picture.
-     Whereas your analysis seems solid and the individual steps you took can easily be followed, I encourage the author(s) to also briefly describe tables/figures to point out what the most relevant information is or what the reader is supposed to see from these tables and figures. However, some important information is missing, and this should be added: you mention that you have a random structure in both regression models. However, nothing is said about the impact of the random effects. Please add. In addition, the model fit of both models needs to be reported.
-     Specific comment to Figure 4/Model 2: Note that if you have an interaction in your model (which you have, namely 3!), you shall interpret the interactions instead of the main effects! This might actually have an effect on your interpretation here, but since this is not given in your results section, it is a bit hard to tell how you interpret the findings. Please add this and also pay attention to how to interpret such regression models. In addition to this, I would suggest considering a step-wise model building process (backward, or forward). You could then, if statistically not significant, remove/or not add

predictors/interactions, rather than keeping them. Take a look, for example at Gries 2021. One possible way could be to start out with the maximum model (all main effects, all two-way interactions) and then to step-wise remove non-significant predictors (either main effects, if not in an interaction, or interactions) via drop1 in R.

-    I am a bit skeptical about the discussion, simply because of the interpretation of Model 2 (see my comment above). Perhaps once this has been taken care off, the discussion needs to be adjusted.

Gries, S. Th. 2021. Statistics for linguists with R. A practical introduction. 3rd edition. De Gruyter Mouton.

_____