**Supporting Information**

**Site-specific data collection information**

US1: A total of 143 participants were recruited and tested at a large, midwestern, public university using the provided materials and protocol for the aural paradigm. Participants were recruited from third-semester basic Spanish language classes. Students enrolled in this course if they had completed the previous semester of the Spanish basic language program or if they had tested into the course based on a placement test. No heritage speakers of Spanish were enrolled in the course, but 44.44% of the students reported being bilingual. Participants were tested in groups in a classroom setting. Testing occurred over two semesters. All participants were compensated with course credit.

US3: A total of 256 participants were recruited and tested at a large, northwestern, public university using the provided materials and protocol for the aural paradigm. Participants were recruited at the start of the second of quarter of the Second-Year Spanish sequence. Students enrolled in this course if they had completed the first quarter of the three-quarter sequence devoted to development of Spanish-language skills with emphasis on diversity of Hispanic cultures or if they had tested into the course based on a placement test. Data from 14 heritage speakers of Spanish were excluded from the analysis. The majority of the remaining 242 participants were raised monolingually, with only 5% of these students reporting being bilingual in English and languages other than Spanish. All participants were tested in groups on the same day during their normal class time in a classroom setting. At the end of the session, after completing the comprehension questions and the proficiency test, the students completed a more detailed language background questionnaire adopted from Mikhaylova (2012). All participants were compensated with course credit.

UK2: A total of 43 participants were recruited and tested at a large, South England, public university using the materials and protocol provided for the aural paradigm. Participants were recruited from stages 3 to 5 and accelerated stages 3 and 4 Spanish classes (i.e., approximately B1 or B2 CEFR level). They were invited to participate by their language instructors and through departmental channels. Most students in this Modern Languages department study at least two foreign languages. 27.91% of the students reported being bilingual. Participants were tested in groups in a computer lab. Testing occurred over two semesters. All participants were compensated monetarily with funds provided by the *Language Learning* Small Grants Research Program.

POL: A total of 59 participants were recruited and tested at university Poland using the materials and protocol provided for the aural paradigm. Participants were recruited from 4 different classes. C1 - a small group consisted of 4 students (Portuguese Studies). Two of them had some contact with Spanish, whereas the other two are 4th graders of Spanish Philology (level C1). So the group is rather mixed. C2 - Spanish Philology students, 2rd graders, after approximately 500 hours of practical Spanish classes. However, there might be some differences in language levels, as some of them had started studying Spanish before they entered the University. C3 - Spanish Philology students, 1st graders, after approximately 200h of practical Spanish classes. Similarly, to the previous group, there might be some mixed levels involved. C4 - The highest group, 3rd grade, Spanish Philology, more than 700 hours of Spanish done. They should be the most equal group, (B2/C1 level). Participants were tested in groups in a classroom setting. Testing occurred over one semester in 4 sessions/each group separately. All participants were compensated with good word and gratefulness. The tests were carried out during their regular classes.

UK1: A total of 62 participants were recruited and tested at two universities in the north of England using the materials and protocol provided for the written paradigm. Participants were recruited from first year undergraduate ‘Spanish BA Honours’ (university major) programs in both universities (for which students had taken ‘A level’ Spanish before coming to University) and also from levels 2, 2+, and 3 from ‘Languages for All’ (institution-wide language programs, i.e. for majors in other subjects). Level 2 is for students with GCSE Spanish or 2-3 years of previous study (approximately equivalent to A2+ of the Common European Framework of Reference (CEFR)); Level 2+ is for students with AS-Level or equivalent or 4-5 years of previous study (approximately A2/B1 of the CEFR); Level 3 is for students with A Level or equivalent (approximately B1 - B2 of the CEFR). Sixteen participants self-reported as being bilingual. Participants were tested in a lab in small groups on the different sites. Testing occurred over two semesters in a number of sessions at each site outside class time. One participant made notes whilst reading the text and was eliminated from the final analysis. All participants were compensated monetarily (£5) with funds provided by the *Language Learning* Small Grants Research Program.

US2: A total of 47 participants were recruited and tested at a private midwestern using the materials and protocol provided for the written paradigm. Participants were recruited from second semester Spanish language classes and were all enrolled to fulfill their foreign language requirement. Classes met three times a week for 50 minutes each session and all four skills were promoted within a communicative framework. Participants were tested in groups in a laboratory. Testing occurred over one semester. All participants were compensated with course credit.

US4: A total of 99 participants were recruited and tested at a southeastern public university using the materials and protocol provided for the written paradigm. Participants were recruited from Beginner Spanish II classes (second-semester Spanish). Students in this class have completed the Beginner Spanish I class, which is a course for false beginners. The great majority of participants only spoke English, and were studying Spanish as their L2. Only seven had a different L2 (Gujurati, Hindu, Greek, Creole, Tagalog, and Albanian). One was multilingual (French, Russian & Greek). Participants signed up for 40-minute blocks and came to the language lab to be tested on their preferred day/time. Testing occurred over two weeks in the Spring semester of 2016. All participants were compensated with extra credit on one of their class compositions.

|  |
| --- |
| Table SI-1*Detailed Participant Information by Modality, Site, and Condition* |
| Site/Condition | Initial *N* | Missing Output File | Excl1 | Excl2 | Semi-Final*N* | Excl3 | Final*N* | Gender*N* female | Age*M* (*SD*) | Numa Lang*M* (*SD*) | Profb*M* (*SD*) | NumChecks/Clicksc*M* (*SD*) |
|  | *Aural paradigm* |  |
| US1-A | 143 | n/a | 2 | 1 | 140 | 14 | 126 | 79 | 20.29 (4.62) | 1.52 (0.76) | 0.39 (0.13) | 9.40(3.61) |
| Control | 37 | n/a | 0 | 1 | 36 | 0 | 36 | 22 | 21.00(6.07) | 1.39(0.49) | 0.39(0.13) | n/a |
| *sol* | 30 | n/a | 0 | 0 | 30 | 1 | 29 | 18 | 21.62(4.57) | 1.38(0.55) | 0.40(0.12) | 8.86(1.64) |
| *la* | 39 | n/a | 0 | 0 | 39 | 3 | 36 | 23 | 18.89(3.59) | 1.67(0.82) | 0.39(0.13) | 9.31(2.25) |
| *-n* | 37 | n/a | 2 | 0 | 35 | 10 | 25 | 16 | 19.76(2.44) | 1.68(1.09) | 0.39(0.12) | 10.16(5.98) |
| US3-A | 241 | n/a | 3 | 0 | 238 | 26 | 212 | 142 | 20.67 (8.25) | 1.08 (0.34) | 0.38 (011) | 9.13d(1.99) |
| Control | 60 | n/a | 0 | 0 | 60 | 0 | 60 | 42 | 20.42(4.14) | 1.08(0.33) | 0.40(0.11) | n/a |
| *sol* | 61 | n/a | 1 | 0 | 60 | 0 | 60 | 48 | 21.05(14.38) | 1.07(0.31) | 0.38(0.10) | 9.17(0.95) |
| *la* | 58 | n/a | 1 | 0 | 57 | 4 | 53 | 34 | 19.94(1.97) | 1.02(0.14) | 0.40(0.11) | 9.77(2.46) |
| *-n* | 62 | n/a | 1 | 0 | 61 | 22 | 39 | 18 | 21.44(4.31) | 1.15(0.53) | 0.35(0.11) | 8.21(2.09) |
| UK2-A | 43 | n/a | 2 | 0 | 41 | 0 | 41 | 34 | 20.73 (6.55) | 1.39 (0.66) | 0.72 (.12) | 10.00(1.81) |
| Control | 11 | n/a | 0 | 0 | 11 | 0 | 11 | 10 | 22.64(11.91) | 1.18(0.58) | 0.66(0.09) | n/a |
| *sol* | 11 | n/a | 2 | 0 | 9 | 0 | 9 | 7 | 19.89(1.74) | 1.56(0.84) | 0.70(0.14) | 9.67(0.47) |
| *la* | 11 | n/a | 0 | 0 | 11 | 0 | 11 | 8 | 19.27(1.22) | 1.55(0.66) | 0.75(0.12) | 10.00(0.74) |
| *-n* | 10 | n/a | 0 | 0 | 10 | 0 | 10 | 9 | 21.00(3.11) | 1.30(0.46) | 0.77(0.07) | 10.30(2.98) |
| POL-A | 59 | n/a | 3 | 0 | 56 | 1 | 55 | 48 | 21.25 (1.70) | 1.00 (0.00) | 0.76 (0.19) | 9.78(1.84) |
| Control | 15 | n/a | 0 | 0 | 15 | 0 | 15 | 15 | 21.27(1.99) | 1.00(0.00) | 0.76(0.20) | n/a |
| *sol* | 15 | n/a | 1 | 0 | 14 | 0 | 14 | 10 | 21.29(1.67) | 1.00(0.00) | 0.76(0.15) | 9.07(1.34) |
| *la* | 15 | n/a | 1 | 0 | 14 | 0 | 14 | 12 | 21.21(1.43) | 1.00(0.00) | 0.78(0.18) | 10.00(1.61) |
| *-n* | 14 | n/a | 1 | 0 | 13 | 1 | 12 | 11 | 20.75(1.59) | 1.00(0.00) | 0.74(0.21) | 10.33(2.29) |
| Total | 486 | n/a | 10 | 1 | 475 | 41 | 434 | 303 | 20.62(6.62) | 1.23(0.56) | 0.47(0.19) | 9.38(2.55) |
|  | *Timed-written paradigm* |  |
| UK1-W | 62 | 0 | 1 | 1 | 60 | 0 | 60 | 48 | 18.45 (2.68) | 1.37 (0.73) | 0.60 (0.16) | 9.24e(1.07) |
| Control | 15 | 0 | 0 | 1 | 14 | 0 | 14 | 11 | 18.36(0.48) | 1.43(0.91) | 0.61(0.19) | n/a |
| *sol* | 15 | 0 | 0 | 0 | 15 | 0 | 15 | 11 | 19.00(1.79) | 1.20(0.40) | 0.54(0.12) | 9.87(0.34) |
| *la* | 16 | 0 | 1 | 0 | 15 | 0 | 15 | 11 | 17.93(4.95) | 1.47(0.72) | 0.60(0.16) | 9.47(0.89) |
| *-n* | 16 | 0 | 0 | 0 | 16 | 0 | 16 | 15 | 18.50(0.61) | 1.38(0.78) | 0.65(0.16) | 8.44(1.17) |
| US2-W | 58 | 3 | 2 | 0 | 53 | 6 | 47 | 20f | 19.17 (0.98) | 1.15 (0.36) | 0.40 (0.11) | 9.09(1.00) |
| Control | 15 | 1 | 0 | 0 | 14 | 0 | 14 | 5 | 19.21(1.15) | 1.07(0.43) | 0.38(0.14) | n/a |
| *sol* | 14 | 0 | 1 | 0 | 13 | 2 | 11 | 5 | 19.5530.99) | 1.18(0.41) | 0.42(0.09) | 9.36(0.77) |
| *la* | 15 | 1 | 1 | 0 | 13 | 2 | 11 | 3 | 19.18(0.84) | 1.18(0.00) | 0.42(0.11) | 9.09(1.17) |
| *-n* | 14 | 1 | 0 | 0 | 13 | 2 | 11 | 7 | 18.73(0.62) | 1.18(0.71) | 0.38(0.10) | 8.82(0.94) |
| US4-W | 98 | 0 | 0 | 0 | 98 | 8 | 90 | 62 | 21.81 (6.09) | 1.13 (0.45) | 0.36 (0.11) | 9.20g(1.04) |
| Control | 25 | 0 | 0 | 0 | 25 | 0 | 25 | 15 | 22.48(6.64) | 1.12(0.43) | 0.36(0.14) | n/a |
| *sol* | 25 | 0 | 0 | 0 | 25 | 2 | 23 | 19 | 21.83(7.15) | 1.22(0.41) | 0.36(0.09) | 9.74(0.44) |
| *la* | 24 | 0 | 0 | 0 | 24 | 0 | 24 | 15 | 20.92(3.75) | 1.00(0.00) | 0.36(0.11) | 9.29(0.98) |
| *-n* | 24 | 0 | 0 | 0 | 24 | 6 | 18 | 13 | 22.06(6.27) | 1.22(0.71) | 0.34(0.10) | 8.39(1.16) |
| Total | 218 | 3 | 3 | 1 | 211 | 14 | 197 | 130 | 20.16(4.66) | 1.21(0.54) | 0.44(0.11) | 9.19(1.04) |
| Overall total | 704 | 3 | 13 | 2 | 686 | 55 | 631 | 433 | 20.48(6.08) | 1.22(0.56) | 0.46(0.19) | 9.32(2.19) |
| *Notes.* Missing Output File = Superlab output file was missing. Excl1 = Participants excluded for not making any checks marks or clicks. Excl2 = Participants excluded for not answering any comprehension questions. Excl3 = Participants excluded for making less than 6 check marks or clicks. aNo difference between conditions for the number of languages was found for any site (*p*s ≥ .219). bNo proficiency differences between conditions was found for any site (*p*s ≥ .091).  cFor the written mode, the number of clicks represents the number of target clicks. dA one-way ANOVA revealed a difference in the number of checks made per condition for US3-A site (*F*(2,149) = 27.697, *p* < .001,  = 0.092). Games-Howell post hoc tests revealed that the –*n* condition made fewer checks than the *sol* and *la* conditions (*p* = .028 and *p* = .004, respectively). eA one-way ANOVA revealed a difference in the number of checks made per condition for UK1-W site (*F*(2,43) = 8.483, *p* < .001,  = 0.324). Games-Howell post hoc tests revealed that the –*n* condition made fewer checks than the *sol* and *la* conditions (*p* < .001 and *p* = .032 respectively). f10 participants did not report gender. gA one-way ANOVA revealed a difference in the number of checks made per condition for US4-W site (*F*(2,62) = 9.365, *p* < .001,  = 0.266). Games-Howell post hoc tests revealed that the –*n* condition made fewer less checks than the *sol* and *la* conditions (*p* = .001 and *p* = .036 respectively). |

|  |
| --- |
| Table SI-2*Descriptive Statistics for Comprehension Accuracy by Mode, Site, and Condition* |
| Site | control | *sol* | *la* | *-n* |
|  | *N* | *M (SD)*[95% CIs] | *N* | *M (SD)*[95% CIs] | *N* | *M (SD)*[95% CIs] | *N* | *M (SD)*[95% CIs] |
|  | *Aural paradigm* |
| US1-A | 36 | 0.28 (0.13)[0.24, 0.33]  | 29 | 0.33 (0.15)[0.27, 0.39] | 36 | 0.33 (0.13)[0.29, 0.38] | 25 | 0.30 (0.14)[0.25, 0.36]  |
| US3-A | 60 | 0.32 (0.15)[0.28, 0.36] | 60 | 0.34 (0.15)[0.30, 0.38] | 53 | 0.35 (0.15)[0.31, 0.39] | 39 | 0.30 (0.15)[0.25, 0.35] |
| UK2-A  | 11 | 0.52 (0.19)[0.39, 0.65] | 9 | 0.47 (0.22)[0.29, 0.64] | 11 | 0.45 (0.13)[0.36, 0.53]  | 10 | 0.36 (0.10)[0.29, 0.43]  |
| POL-A | 15 | 0.41 (0.24)[0.28, 0.54] | 14 | 0.48 (0.20)[0.36, 0.59] | 14 | 0.35 (0.25)[0.21, 0.50] | 12 | 0.23 (0.10)[0.17, 0.29] |
|  | *Timed-written paradigm* |
| UK1-W | 14 | 0.41 (0.15)[0.33, 0.50] | 15 | 0.41 (0.11)[0.35, 0.47] | 15 | 0.45 (0.14)[0.38, 0.53] | 16 | 0.49 (0.16)[0.41, 0.58] |
| US2-W | 14 | 0.46 (0.16)[0.37, 0.55] | 11 | 0.28 (0.12)[0.20, 0.36] | 11 |  0.36 (0.12)[0.27, 0.44] | 11 | 0.35 (0.15)[0.24, 0.45] |
| US4-W | 25 | 0.36 (0.15)[0.30, 0.43] | 23 | 0.38 (0.12)[0.33, 0.44] | 24 | 0.31 (0.13)[0.25, 0.36] | 18 | 0.33 (0.17)[0.25, 0.41] |

|  |
| --- |
| Table SI-3*Effect Sizes for Comprehension Accuracy Comparisons between Experimental and Control Conditions by Mode and Site* |
| Site | *sol* vs control | *la* vs control | *-n* vs control |
|  | *d*[95% CIs] | *d*[95% CIs] | *d*[95% CIs] |
|  | *Original written paradigm* |
| Leow et al. (2008) | 0.28[-0.41, 0.98] | -0.19[-0.95, 0.57] | 0.29[-0.43, 1.02] |
| M-S et al. (2012) | 0.22a[-0.10, 0.54] | 0.28a[-0.06, 0.63] | 0.36\*,a[0.00, 0.71] |
|  | *Aural paradigm* |
| US1-A | 0.35a[-0.15, 0.84] | 0.39a[-0.08, 0.85] | 0.16a[-0.35, 0.67] |
| US3-A | 0.14a,b[-0.22, 0.50] | 0.17a,b[-0.20, 0.54] | -0.15a,b[-0.56, 0.25] |
| UK2-A  | -0.25a[-1.13, 0.64] | -0.45a[-1.30, 0.40] | -1.04\*[-1.95, -0.13] |
| POL-A | 0.32a,b[-0.42, 1.05] | -0.25a[-0.98, 0.49] | -0.94\*[-1.74, -0.14] |
| Meta-analytic effect size | 0.19a[0.00, 0.37] | 0.11a[-0.09, 0.31] | -0.36a[-0.74, 0.03] |
|  | *Timed-written paradigm* |
| UK1-W | -0.05a[-0.78, 0.68] | 0.27a[-0.46, 1.00] | 0.52a[-0.21, 1.25] |
| US2-W | -1.25\*[-2.11, -0.39] | -0.72a[-1.53, 0.09] | -0.73[-1.55, 0.09] |
| US4-W | 0.14a,c[-0.43, 0.71] | -0.40 a,c[-0.97, 0.16] | -0.23a[-0.84, 0.38] |
| Meta-analytic effect size | -0.32[-0.88, 0.25] | -0.28[-0.62, 0.06] | -0.13[-0.59, 0.33] |
| *Note. d* = Cohen’s *d*.\* = 95% CIs do not overlap zero indicating a reliable effect. aThe effect falls within the 95% CI of the respective effect for Leow et al. (2008). bThe effect from an aural replication site falls within the 95% CI of the lead aural site’s respective effect. cThe effect from a timed-written replication site falls within the 95% CI of the lead timed-written site’s respective effect. |