Online supplementary material S1

The Interaction of Central and Peripheral Processing in L2 Handwritten Production: Evidence from Cross-linguistic Variations

Session 1: Evaluations of the English Proficiency and Phonological Abilities

LexTALE test (Lemhöfer & Broersma, 2012) is a rapid, computerized lexical decision task used to estimate proficiency in a second language. Participants were required to judge whether a presented letter string was a real word or not. Scores were calculated based on the number of correct responses. We used the following guidelines threshold for defining high, medium, or low proficiency levels: high proficiency (scores above 75%), medium proficiency (scores between 50% and 75%), and low proficiency (scores below 50%). A minimum score of 72/100 was required to participate in the study.

The Language History Questionnaire (LHQ 3.0; Li et al., 2020) is a comprehensive, self-report questionnaire designed to collect detailed information about participants' linguistic background, proficiency, and immersion in multiple languages. Proficiency and immersion are assessed through multiple items within the LHQ 3.0. For proficiency, participants were asked to rate their abilities in speaking, listening, reading, and writing for the second language on a scale from 1 (very poor) to 7 (very good). We assigned a weight to the self-rated abilities (e.g., speaking: 0.25, listening: 0.25, reading: 0.25, writing: 0.25) and then divided the raw proficiency score by 7 to obtain a normalized proficiency score ranging from 0 to 1. The LexTALE and LHQ-proficiency scores were highly correlated (rho = 0.91, 95% HDI [0.86, 0.94]).

Immersion was assessed through various items that inquire about participants' language exposure and use in different contexts. Information on the second language spoken at home during childhood, the second language used in daily communication, and time spent engaging in activities (e.g., watching TV, reading, or writing) were collected. We summed the weighted values obtained from the three categories above and normalized the scores by dividing each language's total weighted value by the sum of the weighted values. Immersion scores that range from 0 to 1 represent the relative immersion in the second language. The LexTALE and LHQ-immersion scores were also correlated (rho = 0.79, 95% HDI [0.71, 0.85]).

Spoonerism Task (Brunswick et al., 1999) is a phonological awareness assessment that requires

participants to manipulate phonemes while retaining items in working memory. Participants listened to pairs of unrelated words (e.g., basket-lemon) and were asked to swap the initial sounds in each word and read aloud the resulting nonwords (e.g., lasket-bemon). The task includes 12 test items. Following Brunswick et al.'s (1999) methodology, the accuracy score on this task was calculated by tallying the number of item pairs that were accurately produced in their complete form, with the correct pronunciation and in the appropriate order, out of a total of 12 items (see Ben-Yehudah et al., 2019).

Phoneme Deletion Task (M. Wang et al., 2003) assesses participants' ability to manipulate sublexical phonological structures in English through both spoken and written responses. Participants read aloud a word, removed a designated sound within the word (indicated within a phonemic bracket / /), and then read aloud the resulting new word and wrote it down on an answer sheet. The new word must be a real English word, correctly spelled. The task consists of 15 test items. Based on Ben-Yehudah et al. (2019), the accuracy score on the Phoneme Deletion task was determined by computing the number of items that were written and pronounced flawlessly, out of a potential 19 items.

The Nonword Repetition subtest from the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999) measures participants' ability to accurately repeat spoken nonwords. Participants listened to a spoken nonword and then repeated it aloud as accurately as possible. The 18 test items gradually increase in length from one to seven syllables. The accuracy score on the Nonword Repetition subtest was assessed based on standard instructions, measuring the number of items that were pronounced flawlessly out of a total of 18 items.

Table S1

Comparing of the English Proficiency and Phonological Abilities for the Bilingual groups

	Chinese Bilinguals	Spanish Bilinguals	Median	95% HDI	ROPE
LexTALE	79 (4.67)	79.3 (4.15)	0.23	[0.00, 0.46]	0.08
LHQ-proficiency	0.72 (0.09)	0.73 (0.09)	7.47e-03	[0.00, 0.01]	0.04
LHQ-immersion	0.78 (0.1)	0.77 (0.09)	-4.73e-03	[-0.01, 0.00]	0.39
Spoonerism Task	8.95 (1.59)	8.8 (1.42)	-0.14	[-0.33, 0.04]	0.25
Phoneme Deletion	11 (1.87)	11.4 (1.32)	0.4	[-0.08, 0.87]	0.19
The Nonword Repetition	13.6 (2.09)	13.3 (2.21)	-0.30	[-0.68, 0.09]	0.23

Note: values reported are: mean (*sd*).

Session 2: Power Considerations

As stated, one of the hypothetical results we expected to explore was evidence of main or interactive effects between lexical frequency and P-O consistency within each bilingual group. In light of the past research, we assumed the effect size for within-group interactions was moderate according to the criteria set forth for L2 research by Plonsky and Oswald (2014) ($1 - \beta = 0.99$ in Suárez-Coalla et al., 2020, 36 participants; $\eta 2p = 0.12$ in Afonso, Suárez-Coalla, et al., 2015, 40 participants; $d = 0.39$ in Lau, 2021, 32 participants). Meanwhile, following Lau (2021) in a more conservative approach to estimating power, the power to detect the effect estimates as large as 3 standard errors in a full linear mixed-effect model of accuracy data is about 65% (considering the standard errors are around .125 for each fixed effect and .250 for two-way interactions). Over a grand-mean intercept of about 0.3 in this model, a two-way interaction would have to bring accuracy up to 45% or down to 28% to be reasonably detectable. Based on these assumptions, we estimated that we would need at least 40 participants (in each language group) to have an 80% chance of capturing lexical frequency and P-O consistency effects with a type II error rate of 5%. Our prediction related to the potential for a language group effect as a possible mediator of L2 handwritten production is exploratory in nature. We therefore did not base effect size estimates on any parameter estimates related to this between-group predictor, as no existing dataset pointed to whether bilinguals with divergent L1 backgrounds would differ in central and peripheral processes during L2 handwritten production. From a scientifically cautious perspective, the Bayesian approach was applied to incorporate prior beliefs or knowledge about the possible parameters of the model. This method can be particularly beneficial in situations where there is limited prior knowledge of the relationships between the predictors, as it allows for the inclusion of subjective prior information to guide parameter estimation.

Table S2

The List of Experimental Stimuli

Types	Items	Len	NP	NS	OLD20	N	Bigra	Subtlex	TO_FF	TO_FB	TY_FF	TY_FB
Consistent High-frequency	broadcast	9	8	2	3.0	0	49571	4.1	0.9	0.8	0.9	0.8
	collect	7	6	2	2.5	0	40393	4.5	0.5	0.8	0.5	0.8
	finish	7	6	2	1.9	0	40002	5.1	0.6	0.8	0.6	0.8
	gamble	6	6	2	1.6	3	25793	4.2	0.5	1.0	0.5	1.0
	noise	5	4	2	1.7	3	22825	4.8	1.0	1.0	1.0	1.0
	manage	6	5	2	1.8	0	30803	4.9	0.5	1.0	0.5	1.0
	selfish	7	6	2	2.5	0	32511	4.0	0.6	0.7	0.6	0.7
	design	6	5	2	2.0	1	38884	4.8	0.6	0.6	0.6	0.6
Consistent Low-frequency	broiler	7	6	2	1.8	1	44448	1.5	1.0	1.0	1.0	1.0
	collude	7	5	2	2.3	1	37512	2.7	0.5	0.8	0.5	0.8
	fickle	6	5	2	1.8	4	30142	3.1	0.6	0.8	0.6	0.8
	gantry	6	6	2	2.3	2	30180	2.5	0.5	1.0	0.5	1.0
	novice	6	5	2	2.0	1	22548	3.1	0.5	0.8	0.5	0.8
	malice	6	5	2	2.0	0	36683	2.9	0.5	1.0	0.5	1.0
	serif	5	5	2	1.9	0	39679	1.6	0.6	0.7	0.6	0.7
	devote	6	5	2	1.8	2	32968	3.1	0.6	0.6	0.6	0.6
Inconsistent High-frequency	brother	7	6	2	1.9	1	42605	5.0	0.1	0.2	0.1	0.2
	couple	6	5	2	1.8	1	40783	5.4	0.2	0	0.2	0
	finance	7	6	2	2.5	0	50611	4.4	0.1	0.1	0.1	0.1
	garden	6	6	2	1.8	3	33658	5.3	0.1	0.3	0.1	0.3
	nothing	7	5	2	1.8	1	50954	5.6	0.1	0.1	0.1	0.1
	machine	7	5	2	2.2	0	51533	4.3	0.2	0.2	0.2	0.2
	secret	6	6	2	1.9	0	40112	4.9	0.1	0.2	0.1	0.2
	desert	6	6	2	1.9	0	44845	4.3	0.1	0.1	0.1	0.1
Inconsistent Low-frequency	brocade	7	6	2	2.4	0	34736	2.4	0.1	0.2	0.1	0.2
	couture	7	6	2	2.4	0	47068	2.9	0.1	0.3	0.1	0.3
	fiery	5	5	2	2.8	0	25922	3.2	0.1	0.1	0.1	0.1
	galore	6	5	2	2.3	0	38338	3.1	0.1	0.1	0.1	0.1
	nougat	6	5	2	2.4	1	28059	2.4	0.1	0.2	0.1	0.2
	maroon	6	5	2	2.0	0	39647	2.9	0.2	0.2	0.2	0.2
	sewage	6	4	2	2.2	0	31451	3.3	0.1	0.1	0.1	0.1
	demote	6	5	2	1.8	2	35415	2.0	0.1	0.2	0.1	0.2

Note: **Len** = Letter Length; **NP** = Number of Phonemes; **NS** = Number of Syllables; **N** = Coltheart' N;

Bigra = Bigram Frequency; **Subtlex** = Lexical Frequency;

TO_FF = Token Feedforward Consistency; **TO_FB** = Token Feedback Consistency;

TY_FF = Type Feedforward Consistency; **TY_FB** = Type Feedback Consistency.

References

Afonso, O., Suárez-Coalla, P., & Cuetos, F. (2015a). Spelling impairments in spanish dyslexic adults. Frontiers in Psychology, 6, 466.

Lau, D. K.-Y. (2021). The dual-route account of writing-to-dictation in chinese: A short report. Language and Speech, 64(4), 790–803.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of english. Behavior Research Methods, 44, 325–343.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. Bilingualism: Language and Cognition, 23(5), 938-944.

Brunswick, N., McCrory, E., Price, C., Frith, C., & Frith, U. (1999). Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: A search for wernicke's wortschatz? Brain, 122(10), 1901–1917.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. Language Learning, 64(4), 878–912.

Suárez-Coalla, P., Afonso, O., Martínez-García, C., & Cuetos, F. (2020). Dynamics of sentence handwriting in dyslexia: The impact of frequency and consistency. Frontiers in Psychology, 11, 319.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (1999). Comprehensive test of phonological processing: CTOPP. Pro-ed Austin, TX.

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in english word identification: A comparison of korean and chinese english L2 learners. Cognition, 87(2), 129–149.