

Appendix S1

Table 9

Summary of Prior Studies Investigating the Relationship Between Initial Proficiency and L2 Outcomes Abroad

Relationship with L2 outcomes	Study	Program length	N	Proficiency measure(s)	L2 outcome measure(s)
Positive	Golonka (2006)	1 semester	22	*ACTR test *Errors, vocabulary, and self-repair rate in OPI	*OPI
	Davidson (2010)	2-9 months	181	*ACTR grammar test OPI Reading test *Listening test	*OPI Reading test Listening test
	DeKeyser (2010)	6 weeks	16	*MLA test	*Oral interviews
	Leonard & Shea (2017)	3 months	39	Metalinguistic grammar test *DELE vocabulary test	*Oral production task
	Faretta-Stutenberg & Morgan-Short (2018)	1 semester	15	*Elicited Imitation Task *DELE	Oral production task *Grammaticality judgment task
Negative	Brecht, Davidson, & Ginsberg (1995)	4 months	658	ACTR test *OPI **Reading test ***Listening test	*OPI **Reading test ***Listening test
	Llanes & Muñoz (2009)	3-4 weeks	24	*Oral production task	*Oral production task Listening test
	Vande Berg, Connor-Linton, & Paige (2009)	8 weeks-1 year	968	*Simulated OPI	*Simulated OPI
	Baker-Smemoe, Dewey, Bown, & Martinsen (2014)	8-16 weeks	102	*OPI	*OPI

Note. Proficiency and outcome measures that were related are marked with asterisks. Davidson (2010) and Leonard and Shea (2017) also report relationships with initial OPI and oral production performance, respectively. ACTR, American Council of Teachers of Russian; OPI, Oral Proficiency Interview; MLA, Modern Language Association; DELE, Diploma de Español como Lengua Extranjera.

Appendix S2

Table 10
Grammatical Complexity: Robust Coefficients for RQ1 (1,000 Bootstrap Samples)

		<i>B</i>	<i>SE B</i>	Bias	<i>p</i>	<i>BC_a 95% CI B</i>
ΔMLU	<u>Model 1</u>					
	Week 1 performance	-.28	.13	.003	.043	[-.57, .01]
	L2 contact	-.01	.01	.001	.208	[-.03, .01]
	<u>Model 2</u>					
	Week 1 performance	-.39	.13	.001	.008	[-.63, -.12]
	L2 contact	-.01	.01	.001	.401	[-.03, .02]
	EIT	.03	.01	.001	.020	[.01, .05]
ΔDC	<u>Model 1</u>					
	Week 1 performance	-.38	.18	-.03	.047	[-.73, -.12]
	L2 contact	-.003	.002	<.001	.197	[-.007, <.001]
	<u>Model 2</u>					
	Week 1 performance	-.67	.18	-.02	.002	[-1.01, -.37]
	L2 contact	-.002	.002	<.001	.176	[-.005, <.001]
	EIT	.006	.002	<.001	.004	[.002, .009]

Table 11

Grammatical Accuracy: Robust Coefficients for RQ1 (1,000 Bootstrap Samples)

		<i>B</i>	<i>SE B</i>	<i>Bias</i>	<i>p</i>	<i>BC_a 95% CI B</i>	
Δ Subject-verb	<u>Model 1</u>						
	Week 1 performance	-.78	.38	.10	.053	[-1.20, .43]	
	L2 contact	<.001	.001	<.001	.969	[-.001, .001]	
	<u>Model 2</u>						
	Week 1 performance	-1.10	.36	.09	.016	[-1.61, .15]	
	L2 contact	<.001	.001	<.001	.865	[-.001, .001]	
	EIT	-.002	.001	<.001	.031	[-.003, -.001]	
	Δ Number	<u>Model 1</u>					
		Week 1 performance	-.87	.16	.007	.005	[-1.27, -.56]
L2 contact		-.001	.002	<.001	.745	[-.004, .003]	
<u>Model 2</u>							
Week 1 performance		-1.32	.21	.05	.002	[-1.71, -.79]	
L2 contact		-.001	.001	<.001	.597	[-.003, .002]	
EIT		-.007	.002	<.001	.032	[-.01, -.002]	
Δ Gender		<u>Model 1</u>					
		Week 1 performance	-.55	.14	.03	.001	[-.79, .02]
	L2 contact	<.001	.001	<.001	.910	[-.002, .002]	
	<u>Model 2</u>						
	Week 1 performance	-.63	.16	.03	.002	[-.91, .12]	
	L2 contact	<.001	.001	<.001	.929	[-.002, .002]	
	EIT	-.001	.001	<.001	.187	[-.003, .002]	

Table 12

Grammatical Complexity: Robust Coefficients for RQ2 (1,000 Bootstrap Samples)

			<i>B</i>	<i>SE B</i>	Bias	<i>p</i>	<i>BC_a 95% CI B</i>
Δ MLU	<u>Model 1</u>						
		L2 contact	-.01	.01	<.001	.573	[-.03, .01]
	<u>Model 2</u>						
		L2 contact	-.002	.01	.001	.866	[-.02, .02]
		EIT	.016	.01	.001	.168	[-.01, .04]
Δ DC	<u>Model 1</u>						
		L2 contact	-.002	.002	<.001	.341	[-.01, .001]
	<u>Model 2</u>						
		L2 contact	-.001	.002	<.001	.516	[-.004, .002]
		EIT	.003	.002	<.001	.155	[<.001, .01]

Table 13

Grammatical Accuracy: Robust Coefficients for RQ2 (1,000 Bootstrap Samples)

			<i>B</i>	<i>SE B</i>	Bias	<i>p</i>	<i>BC_a 95% CI B</i>
ΔSubject-verb	<u>Model 1</u>						
		L2 contact	-.001	.001	<.001	.125	[-.003, <.001]
	<u>Model 2</u>						
		L2 contact	-.002	.001	<.001	.111	[-.003, <.001]
		EIT	-.001	.001	<.001	.381	[-.002, .001]
ΔNumber	<u>Model 1</u>						
		L2 contact	-.003	.003	<.001	.223	[-.01, .003]
	<u>Model 2</u>						
		L2 contact	-.003	.003	<.001	.256	[-.01, .004]
		EIT	<.001	.002	<.001	.868	[-.01, .004]
ΔGender	<u>Model 1</u>						
		L2 contact	-.001	.002	<.001	.537	[-.01, .003]
	<u>Model 2</u>						
		L2 contact	-.001	.002	<.001	.668	[-.004, .003]
		EIT	.002	.001	<.001	.082	[<.001, .004]

Appendix S3

Galton Squeeze Diagrams: Week 1 Performance (left-axis) and L2 Change (right-axis)

FIGURE 7 HERE

FIGURE 8 HERE

FIGURE 9 HERE

FIGURE 10 HERE

Appendix S4

Table 14

Correlation Matrix (Spearman's): L2 Proficiency and L2 Grammar Performance at Week 1

Measure	1	2	3	4	5
1. EIT	-				
2. MLU	.457*	-			
3. DC	.560**	.761**	-		
4. Subject-verb	-.369*	-.462*	-.441*	-	
5. Number	-.499**	-.400*	-.384*	.391*	-
6. Gender	-.767**	-.295	-.267	.315	.352

Note. MLU, Mean length of AS-unit; DC, Dependent clauses per AS-unit; Subject-verb, Number and Gender represent agreement error ratios.

* $p < .05$; ** $p < .01$