Appendix S1

Table S1. *Background Information of the L2 Speakers*

Time	Variables	Mean (*SD*)	Min-Max
Time 1	Age	26.56 (6.36)	18–49
	Length of Residence (months)	34.78 (33.11)	1–216
	Age of Arrival[a]	22.92 (6.05)	3–40
	Age of Instruction	8.16 (3.88)	2–30
Time 2	Age	27.02 (6.27)	18–49
	Length of Residence (months)	36.19 (34.83)	1–216
	Age of Arrival[a]	23.32 (6.07)	3–40
	Age of Instruction	8.38 (4.02)	3–30

Note. [a]Two participants reported having arrived in English-speaking countries before the age of 14. They had spent 1/5 to 1/3 of their lives (P1: 6.5 years; P2: 4.2 years) in English-speaking countries. These participants remained in the final analyses as they had all received formal education outside of English-speaking countries comprising both form- and meaning-focused instruction and thus likely possessed both explicit and implicit L2 English knowledge

Table S2. *Timeline and Test Sequence*

Setting	Time 1 (January-February, 2019)		Interim	Time 2 (April-May, 2019)	
	Knowledge measures	Min.	Language activity measures	Knowledge measures	Min.
Web	Background questionnaire	15	*Language exposure log*	Background questionnaire	10
	Oral production	10		Oral production	10
	Elicited imitation	15		Elicited imitation	15
Lab	Timed and untimed written GJT	15		Timed and untimed written GJT	15
	MKT	20		MKT	20

Table S3. *Summary of the Descriptive Results*

Constructs	Measures	Findings	Effect Size[CI] [a]
Implicit Knowledge	Elicited Imitation	Improved ($p < .001$)	*d* = 0.35 [0.16, 0.54]
	Oral Production	Improved ($p < .001$)	*d* = 0.33 [0.13, 0.52]
	Timed WGJT	Comparable ($p = 0.127$)	*d* = 0.17 [-0.04, 0.32]
Explicit Knowledge	Untimed WGJT	Improved ($p < .001$)	*d* = 0.85 [0.63, 1.05]
	Metalinguistic Knowledge Test	Improved ($p < .001$)	*d* = 0.90 [0.87, 1.33]

Note. [a] Effect size interpretation is based on Plonsky and Oswald (2014)'s field-specific benchmarks (within-group: small, *d* = 0.60; medium, *d* = 1.00; large, *d* = 1.40)

Table S4. *CFA Model Fit Indices*

	T1	T2
Parameters (*n*)	16	16
χ^2	3.234	7.298
χ^2 *p* (> 0.05)	0.519	0.121
df	4	4
CFI (>= .95)	1.000	0.957
SRMR (<= 0.08)	0.021	0.047
RMSEA	0.000	0.082
RMSEA lower (<= 0.05)	0.000	0.000
RMSEA upper	0.122	0.162

Table S5. *Factor Scores of Explicit and Implicit Knowledge at T1 and T2*

	Mean	*SD*	Min	Max	Skewness	Kurtosis
Implicit T1	0.000	0.084	-0.303	0.173	-0.582	0.767
Explicit T1	0.000	0.167	-0.402	0.338	-0.264	-0.515
Implicit T2	0.000	0.078	-0.208	0.178	-0.187	-0.170
Explicit T2	0.000	0.097	-0.367	0.193	-0.748	0.965
Meaning	0.000	1.000	-1.742	6.435	2.375	12.526
Language	0.000	1.000	-0.970	7.387	3.911	23.736

Note. Implicit = implicit knowledge; Explicit = explicit knowledge; T1 = Time 1; T2 = Time 2; Meaning = meaning-focused activity; Language = language-focused activity.

Appendix S2

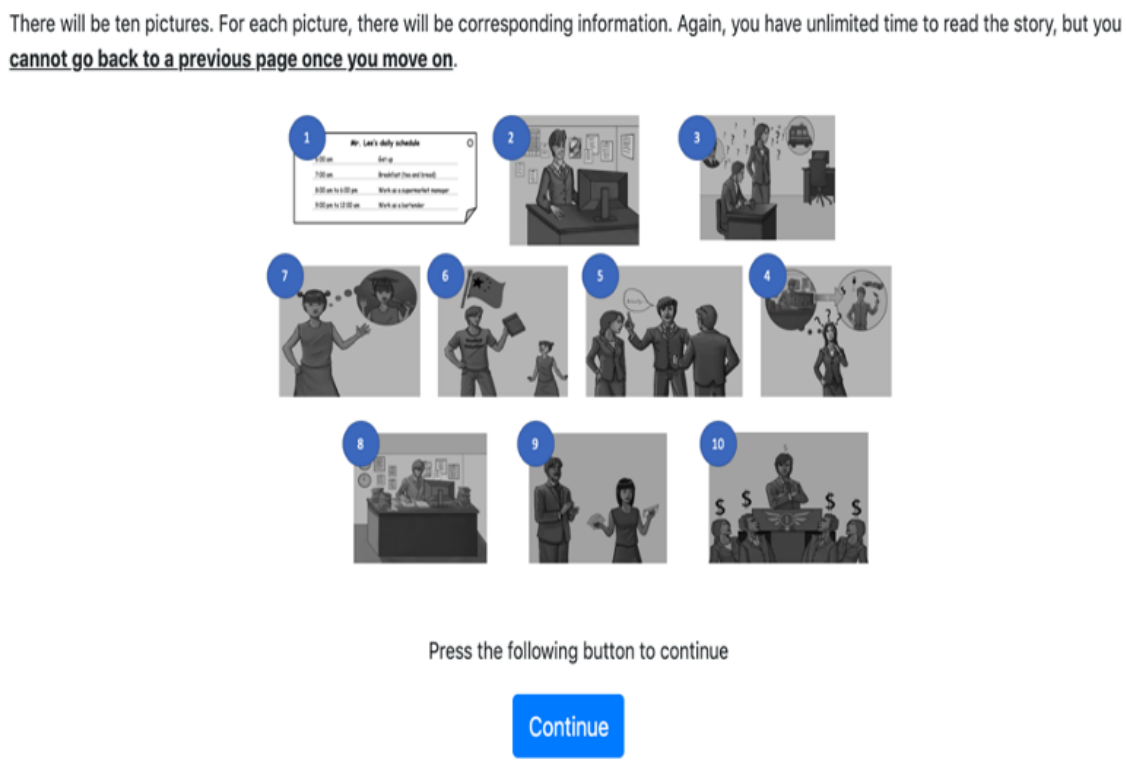

Figure S**1.** Web-based instructions in the oral production task

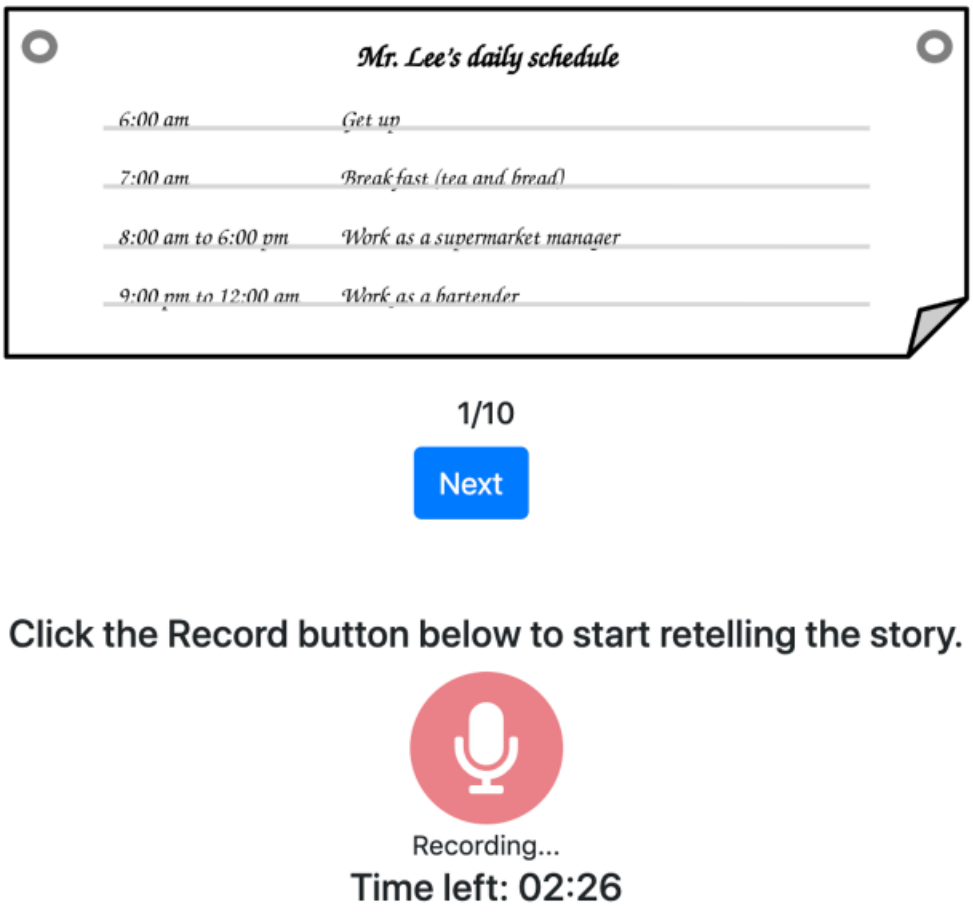

Figure S2. The oral production task during retelling with picture prompts and a progress indicator

Appendix S3. Missing data

As in most longitudinal studies, there were multiple missing values. In particular, there were missing test scores within and across sessions. For instance, participants completed the web-based tasks but did not come to the lab session to complete the remaining tests (missing test data within sessions) and there were participant dropouts (missing data across sessions). As seen in Table S4, the missingness of individual measures of 149 participants, at a test score level, ranged from 0.67% to 24.83%, with an average of 14.83%, which is not uncommon in longitudinal studies (see Schoonen, van Gelderen, Stoel, Hulstijn & de Glopper, 2011). The highest percentage of missing data came from oral productions at T1 and T2, with 14 (out of 26) and 6 (out of 37) missing data due to technical errors (i.e., poor quality or corrupted file).

Table **S4**.1 Missing Data of 149 Participants in T1 And T2

Measures	Time	Missing (n)	Missing (%)
EI	1	10	0.67%
OP	1	26	17.45%
TGJT	1	8	5.37%
UGJT	1	8	5.37%
MKT	1	8	5.37%

EI	2	34	22.82%
OP	2	37	24.83%
TGJT	2	31	20.81%
UGJT	2	31	20.81%
MKT	2	28	18.79%

Note. EI = elicited imitation; OP = oral production; TGJT = timed written grammaticality judgment task; UGJT = untimed written grammaticality judgment task; MKT = metalinguistic knowledge test

In handling missing values at the test levels, we used a model-based approach (i.e., full-information maximum likelihood estimation) that produces parameter estimates of models in the presence of missing data. This meant all 149 participants' data from T1 could be included without needing to remove the 28 dropouts. An important assumption made in this approach is that the data are missing at random (MAR) or completely at random (MCAR). While MCAR is not likely to be met with longitudinal data since missingness is typically correlated with proficiency, MAR can be met when the missing data can be predicted from observed data (Little & Rubin, 1987). In our data, missing patterns are partially traceable by the types of missing tests. The left panel of Figure S4 visualizes the proportion of missing values across two-time points and the right panel provides combinations of missing data patterns. The red-filled

squares represent missing values and the numbers on the right indicate participants who fall in a given category. As visualized in the right panel of Figure S4, seven participants without TGJT, UGJT, and MKT scores (the three lab-based tests) at T1 dropped out of the experiment; 14 participants without OP scores at T1 decided to discontinue the study. In addition to this, all variables in this data set are sufficiently correlated, which is an essential element for drawing reliable estimates of missing scores. With this evidence in mind, we discerned the data to be MAR and carried out a full-information maximum likelihood estimation in evaluating different models.

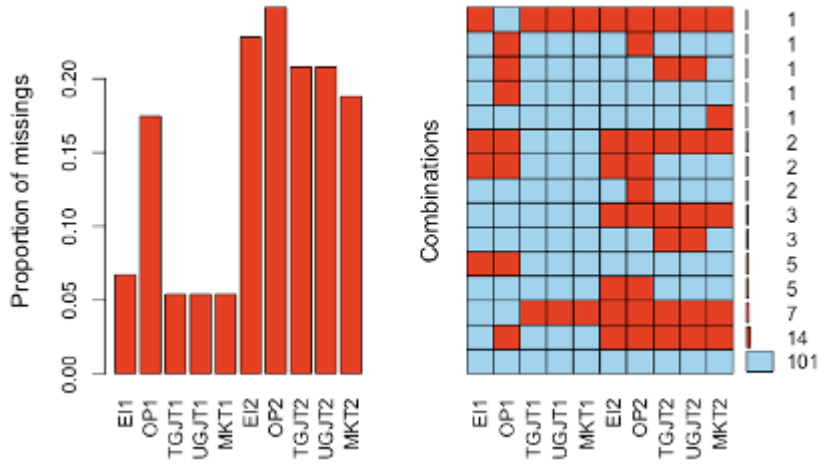

Figure S4. (left panel) Proportion of missing values across two-time points; (right panel) Combinations of missing data patterns. Numbers on the right indicate participants who fall in a given category. Red-filled squares represent missing values