Additional Supplementary Material: Partial Analyses with 119 Participants (Outlier Candidate Removed)

1. Intelligibility (HI recordings) Results

As shown in Table 1, Model 6 was selected as the optimal random effects structure. Wald *z* tests for the fixed effects (Table 2) indicated significant main effects of L1 and PreTest. No issues were detected in any of the diagnostic tests conducted using the DHARMa package (v0.4.7; Hartig, 2024) for Model 6. The conditional R^2 and marginal R^2 were 0.595 and 0.086, respectively.

Table 1: *Random Effects Model Comparison for Intelligibility Scores (HI)*

Sample size	Total observations = 5593 Subjects = 119; Items = 47									
Selection process	**Model No.**	**Random effects**		**Model fit**			**Singular ?**	**LRT test**			
		Subject	**Items**	**AIC**	**BIC**	**logLik**		**Reference model**	**df**	**χ^2**	***p***
1. Building a maximal model											
	1	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 | item)	3573	3765.2	-1757.5	Yes	-	-	-	-
2. Building a ZCP model											
	2	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 || item)	3572.2	3665	-1772.1	Yes	-	-	-	-
3. Simplifying the ZCP model with LRT											
	3	(1 | sub)	(1 + Accent + Caption + L1 + Accent:L1 || item)	3570.2	3656.3	-1772.1	Yes	-	-	-	-
	4	(1 | sub)	(1 + Accent + L1 + Accent:L1 || item)	3568.2	3647.7	-1772.1	No	-	-	-	-
	5	(1 | sub)	(1 + L1 + Accent:L1 || item)	3610	3683	-1794	No	4	1	0.09	.767
	6	(1 | sub)	(1 + L1 || item)	3564.4	3630.7	-1772.2	No	5	1	0.18	.669
	7	(1 | sub)	(1 | item)	3570.7	3630.4	-1776.3	No	6	1	8.26	.004
4. Comparing the selected model in LRT process with the model with correlation parameters											
	8	(1 | sub)	(1 + L1 | item)	3566.3	3639.2	-1772.1	No	6	1	0.15	.700

Table 2: *Results of Model 6 for Intelligibility Scores (HI)*

	Fixed effects				Random effects	
					By Subject	By Items
Parameters	Estimate	*SE*	*z*	*p*	*SD*	*SD*
Intercept	2.61	0.30	8.81	< .001	0.68	1.92
Accent	-0.02	0.16	-0.10	.922	-	-
Caption	-0.16	0.15	-1.05	.296	-	-
L1	-0.41	0.18	-2.33	.020	-	0.50
PreTest (covariate)	0.81	0.08	10.04	< .001	-	-
Accent x Caption	0.01	0.31	0.03	.977	-	-
Accent x L1	0.28	0.31	0.89	.372	-	-

2. Intelligibility (LI recordings) Results

The random-effects structure selection indicated that Model 5 was optimal (see Table 3). Wald z tests for the fixed effects (Table 4) showed significant effects of Caption, L1, and PreTest. However, the coefficient for Caption was negative, indicating a negative effect. Model 5 diagnostics revealed no notable issues. The conditional and marginal R^2 values were 0.624 and 0.067, respectively.

Table 3: *Random Effects Model Comparison for Intelligibility Scores (LI)*

Sample size	Total observations = 5712 Subjects = 119; Items = 48									
Selection process	**Model No.**	**Random effects**		**Model fit**			**Singular ?**	**LRT test**			
		Subject	**Items**	**AIC**	**BIC**	**logLik**		**Reference model**	**df**	**χ^2**	***p***
1. Building a maximal model											
	1	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 | item)	4954.2	5147	-2448.1	Yes	-	-	-	-
2. Building a ZCP model											
	2	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 || item)	4933.3	5026.4	-2452.6	Yes	-	-	-	-
3. Simplifying the ZCP model with LRT											
	3	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption || item)	4931.3	5017.7	-2452.6	No	-	-	-	-
	4	(1 | sub)	(1 + Accent + L1 + Accent:Caption || item)	4929.3	5009.1	-2452.7	No	3	1	0.06	.809
	5	(1 | sub)	(1 + L1 + Accent:Caption || item)	4927.8	5001	-2452.9	No	4	1	0.51	.475
	6	(1 | sub)	(1 + L1 || item)	4927.6	4994.1	-2453.8	No	5	1	1.77	.184
4. Comparing the selected model in LRT process with the model with correlation parameters											
	7	(1 | sub)	(1 + L1 + Accent:Caption | item)	4932.6	5025.7	-2452.3	No	5	3	1.24	.743

Table 4: *Results of Model 5 for Intelligibility Scores (LI)*

	Fixed effects				Random effects	
					By subject	By items
Parameters	Estimate	*SE*	*z*	*p*	*SD*	*SD*
Intercept	1.08	0.31	3.44	< .001	0.68	2.10
Accent	0.21	0.14	1.48	.139	-	-
Caption	-0.32	0.14	-2.21	.027	-	-
L1	-0.52	0.17	-3.06	.002	-	0.54
PreTest (covariate)	0.63	0.07	8.54	< .001	-	-
Accent x Caption	0.44	0.30	1.47	.143	-	0.53
Accent x L1	-0.11	0.29	-0.38	.702	-	-

3. Comprehensibility (HI recordings) Results

The optimal random effects structure was Model 6 (see Table 5). Only PreTest was statistically significant in the fixed effects *t* tests (Table 6). Model diagnostics identified a few issues in the full model; however, visual inspection suggested that these were minor. First, the KS test was significant ($p = .008$); nevertheless, the QQ plot showed near-linear agreement with the 45° reference line. Second, although the combined adjusted quantile test was significant, deviations in the residuals-versus-fitted plot were not substantial. The conditional and marginal R^2 values were 0.460 and 0.235, respectively.

Table 5: *Random Effects Model Comparison for Comprehensibility Ratings (HI)*

Sample size	Total observations = 1309 Subjects = 119; Items = 11									
Selection process	**Model No.**	**Random effects**		**Model fit**			**Singular ?**	**LRT test**			
		Subject	**Items**	**AIC**	**BIC**	**logLik**		**Reference model**	**df**	**χ^2**	***p***
1. Building a maximal model											
	1	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 | item)	5139.1	5294.4	-2539.5	Yes	-	-	-	-
2. Building a ZCP model											
	2	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 || item)	5121.4	5199.1	-2545.7	Yes	-	-	-	-
3. Simplifying the ZCP model with LRT											
	3	(1 | sub)	(1 + Accent + Caption + L1 + Accent:L1 || item)	5119.4	5191.9	-2545.7	Yes	-	-	-	-
	4	(1 | sub)	(1 + Accent + Caption + L1 || item)	5117.4	5184.7	-2545.7	No	-	-	-	-
	5	(1 | sub)	(1 + Accent + L1 || item)	5115.6	5177.8	-2545.8	No	4	1	0.20	.657
	6	(1 | sub)	(1 + L1 || item)	5114.4	5171.4	-2546.2	No	5	1	0.80	.372
	7	(1 | sub)	(1 | item)	5128.5	5180.3	-2554.2	No	6	1	15.26	< .001
4. Comparing the selected model in LRT process with the model with correlation parameters											
	8	(1 | sub)	(1 + L1 | item)	5116.0	5178.1	-2546.0	No	6	1	0.46	.498

Table 6: *Results of Model 6 for Comprehensibility Ratings (HI)*

	Fixed effects				Random effects	
					By subject	By items
Parameters	Estimate	*SE*	*t*	*p*	*SD*	*SD*
Intercept	5.55	0.26	21.04	< .001	0.58	0.84
Accent	0.01	0.14	0.05	.962	-	-
Caption	-0.06	0.14	-0.44	.663	-	-
L1	-0.10	0.21	-0.48	.634	-	0.51
PreTest (covariate)	1.05	0.07	14.48	< .001	-	-
Accent x Caption	0.37	0.28	1.32	.191	-	-
Accent x L1	0.32	0.28	1.13	.260	-	-

4. Comprehensibility (LI recordings) Results

Based on LRTs, Model 5 was selected as the optimal random-effects structure (see Table 7). As shown in Table 8, t-tests for the fixed effects indicated that Accent and PreTest were statistically significant. Model diagnostics for Model 5 revealed no notable issues. The conditional and marginal R^2 values were 0.497 and 0.218, respectively.

Table 7: *Random Effects Model Comparison for Comprehensibility Ratings (LI)*

Sample size	Total observations = 1309 Subjects = 119; Items = 11									
Selection process	**Model No.**	**Random effects**		**Model fit**			**Singular ?**	**LRT test**			
		Subject	**Items**	**AIC**	**BIC**	**logLik**		**Reference model**	**df**	**χ^2**	***p***
1. Building a maximal model											
	1	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 | item)	5123.7	5279.0	-2531.9	Yes	-	-	-	-
2. Building a ZCP model											
	2	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption + Accent:L1 || item)	5103.2	5180.9	-2536.6	Yes	-	-	-	-
3. Simplifying the ZCP model with LRT											
	3	(1 | sub)	(1 + Accent + Caption + L1 + Accent:Caption || item)	5101.2	5173.7	-2536.6	Yes	-	-	-	-
	4	(1 | sub)	(1 + Accent + L1 + Accent:Caption || item)	5099.2	5166.5	-2536.6	No	-	-	-	-
	5	(1 | sub)	(1 + L1 + Accent:Caption || item)	5097.3	5159.4	-2536.6	No	4	1	0.05	.817
	6	(1 | sub)	(1 + Accent:Caption || item)	5099.0	5155.9	-2538.5	No	5	1	3.71	.054
4. Comparing the selected model in LRT process with the model with correlation parameters											
	7	(1 | sub)	(1 + L1 + Accent:Caption | item)	5099.1	5176.8	-2534.6	Yes	-	-	-	-

Table 8: *Results of Model 5 for Comprehensibility Ratings (HI)*

	Fixed effects				Random effects	
					By subject	By items
Parameters	Estimate	*SE*	*t*	*p*	*SD*	*SD*
Intercept	4.45	0.31	14.36	< .001	0.61	1.00
Accent	0.28	0.14	1.99	.049	-	-
Caption	-0.15	0.14	-1.08	.283	-	-
L1	-0.26	0.17	-1.55	.132	-	0.31
PreTest (covariate)	1.01	0.07	14.02	< .001	-	-
Accent x Caption	0.40	0.30	1.33	.192	-	0.37
Accent x L1	-0.24	0.28	-0.84	.400	-	-