**Supplementary Materials**

**Appendix S1**

The present study’s HL learners’ performance on Solon et al.’s (2019) 36-item Spanish EIT were examined using the same procedure described above for the 30-item EIT. Similar to the results for the 30-item EIT, results from Rasch modeling with the 36-item EIT demonstrated generally good fit, with 83% of the items (*k* = 30/36) falling within the 0.5 to 1.5 infit and outfit MNSQ range of fit indices. Six items fell outside the acceptable threshold and thus are considered misfitting (see Table S1): Five items (Items 1, 4, 18, 30, 35) displayed MNSQ statistics higher than 1.5, indicating that they underfitted the model; one item (Item 10) displayed MNSQ statistics below 0.5, suggesting that it overfitted the model. Five of the six misfitting items (Items 1, 4, 10, 18, and 30) were also misfitting on the 30-item EIT, and one new item (Item 35) was also found to be misfitting. This finding suggests that most, but not all, of the six more challenging items added by Solon et al. (2019) are functioning appropriately for HL learners.

Table S1

Rasch fit statistics for the 36-item EIT

|  |  |  |  |
| --- | --- | --- | --- |
|  | Measure | Infit | Outfit |
|  | Mean | *SE* | MNSQ | ZSTD | MNSQ | ZSTD |
| Item 1 | -2.66 | .28 | 1.36 | 1.20 | **2.57** | 1.66 |
| Item 2 | -4.13 | .35 | .91 | -.16 | 1.14 | .50 |
| Item 3 | -1.89 | .25 | 1.09 | .45 | 1.10 | .36 |
| Item 4 | -3.09 | .30 | **1.97** | 2.59 | 1.39 | .69 |
| Item 5 | -1.05 | .21 | 1.07 | .41 | .94 | -.02 |
| Item 6 | -.66 | .20 | 1.06 | .37 | .83 | -.43 |
| Item 7 | -.22 | .20 | .99 | .00 | .82 | .83 |
| Item 8 | -.18 | .19 | 1.23 | 1.20 | 1.20 | .76 |
| Item 9 | -1.50 | .23 | 1.16 | .74 | 1.07 | .30 |
| Item 10 | -1.05 | .21 | **.49** | -3.06 | **.36** | -2.02 |
| Item 11 | .43 | .19 | .96 | -.17 | .89 | -.43 |
| Item 12 | .08 | .19 | 1.02 | .19 | .85 | -.54 |
| Item 13 | .22 | .19 | .79 | -1.15 | 1.32 | 1.26 |
| Item 14 | .67 | .18 | .98 | -.08 | .99 | .02 |
| Item 15 | -1.10 | .22 | .90 | -.45 | .69 | -.69 |
| Item 16 | -.11 | .19 | .80 | -1.01 | .81 | -.68 |
| Item 17 | .26 | .19 | 1.02 | .16 | 1.11 | .53 |
| Item 18 | .00 | .19 | 1.03 | .22 | **2.30** | 3.76 |
| Item 19 | 1.37 | .18 | 1.04 | .28 | 1.01 | .13 |
| Item 20 | .11 | .19 | .69 | -1.81 | .83 | -.61 |
| Item 21 | .94 | .18 | 1.28 | 1.52 | 1.17 | .86 |
| Item 22 | .29 | .19 | .99 | -.02 | .98 | -.01 |
| Item 23 | .00 | .19 | 1.00 | .05 | .86 | -.46 |
| Item 24 | 1.27 | .18 | .96 | -.93 | .87 | -.60 |
| Item 25 | .52 | .20 | .90 | -.21 | .98 | -.02 |
| Item 26 | 1.07 | .18 | .75 | -1.51 | .83 | -.87 |
| Item 27 | -.42 | .20 | .87 | -.62 | .88 | -.50 |
| Item 28 | .43 | .19 | .76 | -1.38 | .64 | -1.74 |
| Item 29 | .40 | .19 | .69 | -1.86 | .67 | -1.54 |
| Item 30 | -1.01 | .21 | 1.29 | 1.39 | **1.75** | 1.68 |
| Item 31 | .91 | .18 | .58 | -2.84 | .53 | -2.79 |
| Item 32 | 1.50 | .18 | .97 | -.14 | .95 | -.21 |
| Item 33 | 1.44 | .18 | .70 | -1.91 | .66 | -1.97 |
| Item 34 | 2.96 | .20 | 1.37 | 1.80 | 1.34 | 1.43 |
| Item 35 | 1.84 | .18 | 1.38 | 1.95 | **1.63** | 2.85 |
| Item 36 | 2.40 | .19 | .90 | -.48 | .91 | -.37 |

An item-person analysis, presented in the Wright map in Figure S1, revealed that the HL learners’ ability covered a range of 9.49 logits, and item difficulty covered a range of 7.09 logits. These findings suggest that the range of item difficulty for the 36-item EIT still did not map adequately with person ability, although they were better matched for the 36-item EIT than for the 30-item EIT. The person ability mean (1.62) was above item mean difficulty (0.00), indicating, again, that the test was easy for many HL learners despite the six additional items added to the test, although the difference in means was smaller than with the 30-item EIT.



Figure S1. Wright map of person ability and item difficulty (*N* = 63)

As with the 30-item EIT, the 36-item versiondemonstrated excellent reliability. Both the item and person reliability coefficients were 0.98, meaning that only 2% of the variability in item and person measures was attributable to error. Item and person separation indices were 6.45 and 6.83, respectively, suggesting that the EIT consistently discriminated approximately seven statistically distinct strata of performance in persons and six difficulty levels in items. These findings suggest that the 36-item EIT better discriminates across persons and items as compared to the 30-item EIT.

Finally, a DIF analysis of these six additional items (see Table S2) revealed that five of the six new items functioned similarly for the present study’s HL learners as for Solon et al.’s (2019) L2 learners. Item 31, however, favored HL learners. Similar to items previously described to favor HL learners, this item (*Son ellas las que acaban de decorar la sala de espera* “They [fem.] are the ones who just finished decorating the waiting room”) likewise includes non-canonical word order, pronominal reduplication, and lexical items (e.g., *sala de espera* “waiting room”) that seemed more problematic for L2 learners than for HL learners. Thus, as with Solon et al.’s (2019) advanced L2 learner sample, the 36-item EIT appears to present some benefits for HL proficiency assessment, at least with a sample that includes high ability HL learners.

Table S2.

DIF analysis of six additional EIT items for HL learners and L2 learners (*N* = 151)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item | Class A | DIF | DIF *SE* | Class B | DIF | DIF *SE* | DIF Contrast | Welch *t* | *df* | *p* |
| 31 | HL | .82 | .17 | L2 | 1.64 | .16 | **-.82** | -3.53 | 139 | .00 |
| 32 | HL | 1.35 | .17 | L2 | 1.21 | .15 | .14 | .62 | 138 | .54 |
| 33 | HL | 1.29 | .17 | L2 | 1.64 | .16 | -.35 | -1.51 | 139 | .13 |
| 34 | HL | 2.62 | .18 | L2 | 2.36 | .16 | .26 | 1.07 | 138 | .28 |
| 35 | HL | 1.65 | .17 | L2 | 1.19 | .15 | .46 | 1.99 | 138 | .05 |
| 36 | HL | 2.14 | .18 | L2 | 1.67 | .16 | .47 | 1.97 | 137 | .05 |

*Note.* DIF contrast greater than .64 is indicated with bolding.