**Appendix A**

*Coding Scheme*

|  |
| --- |
| **Basic information** |
|  | Author | Open |
|  | Year | Open |
|  | Journal | Open |
|  | Document type | 1 = journal article2 = book chapter3 = dissertation/thesis4 = conference paper/poster5 = other |
|  | Peer reviewed | 0 = no1 = yes |
|  | Funded research | 0 = no1 = yes |
| **Design/Study background information** |
|  | One-shot or developmental study? | 0 = one shot1 = developmental |
|  | If developmental, was EIT used to measure effectiveness of a treatment? | 0 = no1 = yes |
|  | Learning context | 0=classroom1=learned for study2=non-classroom |
|  | Comparison of proficiency (Data collected across proficiency levels) | 0=no1=yes |
|  | Participants mean age | Open |
|  | Instructional institution | 1 = elementary/primary2 = secondary3 = college/university4 = language institute5 = not classroom learners99 = not reported |
|  | L1 control group included in study | 0 = no1 = yes |
|  | L2 context | 1 = second language2 = foreign language3 = artificial language |
|  | Research questions clearly stated | 0 = no1 = yes |
| **Participant features** |
|  | Mean participant age | Open |
|  | L1 | Open |
|  | Were L1s compared in the study? | 0 = no1 = yes |
|  | Target L2 | Open |
|  | If target language is an L3, name L2 | Open |
|  | L2 proficiency level | 1 = Beginner2 = Intermediate3 = Advanced4 = Uninterpretable  |
|  | L2 proficiency justification | 0 = no justification1 = assumed from institutional level2 = standardized proficiency test (commercially available and developed by testing agencies)3 = self-reported 4 = other |
|  | Sampling | 1 = random sampling2 = convenience sampling3 = purposive sampling4 = population sampling99 = not reported |
|  | Participant N | Open |
|  | Native speakers control group | 0 = no1 = yes |
|  | Size of native speaker control group | Open |
| **Instrument features** |
|  | Test name | Open |
|  | Test development | 1=researcher-made (created by the author(s) of the study or adapted from a previous study)2=standardized (commercially available and developed by testing agencies)3=classroom-based (developed by a language instructor for classroom purposes) |
|  | Number of target items | Open |
|  | Break | 0 = no1 = yes |
|  | Format | 0 = face-to-face1 = computer-based |
|  | Delayed responses |  0 = no1 = yes |
|  | Method of delay | 1 = silent2 = picture3 = question4 = other |
|  | Construct measured | 0 = global1 = specific |
|  | Stimuli length maximum | Open |
|  | Stimuli length range | Open |
|  | Timed | 0 = no1 = yes |
|  | Ungrammatical stimuli included | 0 = no1 = yes |
|  | Counter-balanced forms | 0 = no1 = yes |
|  | Item randomization | 0 = no1 = yes |
|  | Scale type | 1 = holistic2 = analytic |
|  | Scale | 1 = binary2 = ordinal3 = interval4 = mixed |
|  | Methods used to control difficulty | 0 = no1 = yes |
|  | How was difficulty controlled? | 1 = sentence length2 = lexical3 = syntactic complexity4 = grammatical structure5 = other6 = multiple0 = not controlled |
|  | Were participants asked a comprehension questions after EIT? | 0 = no1 = yes |
| **Reliability, validity, and transparency** |
|  | Was the reliability for the task provided? | 0 = no1 = yes |
|  | Reliability index, if reported. | 0 = none1 = Cronbach's alpha2 = KR-203 = correlation with other test4 = other5 = multiple |
|  | Reliability as reported | Open |
|  | Was another instrument used to triangulate data from the EIT? | 0 = no1 = yes |
|  | If so, what instrument? | Open |
|  | Did the study compare results of multiple EITs? | 0 = no1 = yes |
|  | Test purpose | 1 = admission2 = placement3 = classroom4 = proficiency5 = diagnostic6 = treatment effectiveness7 = non-specified8 = other |
| **Statistical analyses and results** |
|  | What correlations are reported in the study? | 1 = Pearson’s *r*2 = Spearman’s *ρ*3 = other |
|  | Correlation value reported between EIT and another non-standardized test | Open |
|  | Correlation value reported between EIT and end-of-year achievement test | Open |
|  | Correlation value reported between EIT and self-report | Open |
|  | Correlation value reported between EIT and self-reported listening | Open |
|  | Correlation value reported between EIT and self-reported speaking | Open |
|  | Correlation value reported between EIT and self-reported reading | Open |
|  | Correlation value reported between EIT and self-reported writing | Open |
|  | Correlation value reported between EIT and overall teacher evaluation | Open |
|  | Correlation value reported between EIT and teacher evaluation (listening) |  |
|  | Correlation value reported between EIT and teacher evaluation (reading) |  |
|  | Correlation value reported between EIT and teacher evaluation (writing) |  |
|  | Correlation value reported between EIT and teacher evaluation (speaking) |  |
|  | Correlation value reported between EIT and another standardized test | Open |
|  | Correlation value reported between EIT and speaking test/test section | Open |
|  | Correlation value reported between EIT and listening test/test section |  Open |
|  | Correlation value reported between EIT and reading test/test section | Open |
|  | Correlation value reported between EIT and writing test/test section | Open |
|  | Post hoc tests reported | 0 = no1 = yes |
|  | Checked assumptions reported | 0 = no1 = yes |
|  | Inter-rater reliability reported | 0 = no1 = yes |
|  | Inter-rater reliability type | 1 = Cohen's kappa2 = percentage agreement3 = Cronbach's alpha |
|  | Inter-rater reliability value | (numerical value) |
|  | Visuals included | 0 = no1 = yes |
|  | Effect sizes reported (apart from correlations) | 0 = no1 = yes |
|  | Effect size type | 1 = Cohen’s *d* (between means)2 = Hedge’s *g* (between means)3 = eta-squared4 = partial eta-squared5 = Cramer's V (odds ratio)6 = Cohen's f2 |
|  | Effect size value | (numerical input) |