**Supplementary Materials**

**S1: Number of participants with valid data in each measure of individual differences.** The table below shows the number of participants with valid data in the different measures of individual differences (out of a total of N=661), i.e., the following tasks: spelling, vocabulary, motivation, TOWRE, LexTALE, and CFT. Missing values were due to various reasons. In the spelling, vocabulary, and motivation tasks, missing values were spurious (e.g., due to technical and administration errors, time running out for the experimental session) and did not affect the data from a given site substantially. However, for LexTALE, TOWRE and CFT, there were specific reasons that have led to missing values from a full site (e.g., partners in specific sites opting to opt out of the administration of the CFT and TOWRE due to potential copyright restrictions in their jurisdiction). These cases are specified with notes in the table below.

|  |  |
| --- | --- |
| **Measure** | **N valid** |
| Spelling | 648 |
| Vocabulary | 631 |
| Motivation | 616 |
| LexTALE | 563a |
| TOWRE | 558a,b |
| CFT | 560a,b |

Notes: aNo LexTALE, TOWRE, and CFT data was recorded in the Brazilian Portuguese site due to programmatic error; bNo data was collected in TOWRE and CFT in Serbia due to copyright restrictions.

**S2: Specifications of the eye-tracking laboratory setup.** Information regardingapparatus and reading task's settings.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Country** | **Institute** | **L1 of sample** | **Eye-tracker** | **Font point size** | **Distance from screen, in cm** | **Screen type** | **Screen resolution** | **Screen size, in inch** |
| Brazil | Federal Universities of Ceara and Minas Gerais | Portuguese | EyeLink 1000 | 20 | 75 | IPS-LCD | 1920x1080 | 19 |
| China | University of Science and Technology Beijing | Mandarin | Eyelink-1000+ | 20 | 60 | DELL P1917S | 1024 x 768 | 19 |
| Chile | Pontificia Universidad Católica de Valparaíso | Spanish | EyeLink Portable Duo | 20 | 70 | Notebook Gamer Rog Zephyrus M16 | 1920 x 1080 | 16 |
| Denmark | Aalborg University | Danish | EyeLink Portable Duo | 20 | 95 | Dell S2421HGF | 1920 x 1080 | 24 |
| Germany | University of Potsdam | German | EyeLink Portable Duo | 20 | 91 | ASUS ROG Strix XG259QN | 1920 x 1080 | 24 |
| Iceland | University of Iceland | Icelandic | EyeLink 1000+ | 20 | 96 | BenQ, XL2411Z | 1920 x 1080 | 24 |
| India | Indian Institute of Technology Kanpur | Hindi | EyeLink 1000+ | 20 | 70 | BenQ-XL2430T | 1920 x 1080 | 24 |
| India | International Institute of Information Technology Hyderabad | Hindi | EyeLink 1000+ | 20 | 90 cm | Dell S2415Hb | 1920 X 1080 | 24 |
| Russia | Higher School of Economics | Russian | EyeLink 1000+ | 22 | 90 | ASUS VG248QE | 1920 x 1080 | 24 |
| Serbia | Universities of Belgrade and Novi Sad | Serbian | EyeLink II | 22 | 90 | ViewSonic Graphics Display G90FB | 1280 x 1024 | 19 |
| Spain | Basque Center on Cognition, Brain and Language | Basque | Eyelink 1000 | 20 | 60 | Viewsonic CRT | 1600 x 1024 | 19 |
| Switzerland | University of Zurich | German | EyeLink Portable Duo | 10 | 60 | ASUS OMEN | 1280 x 1024 | 24 |
| UK | University of Southampton | English | Eyelink 1000 | 20 | 70 | ASUS VG248QE | 1920x1080 | 24 |
| \*Norway | University of Oslo | Norwegian | EyeLink 1000+ | 20 | 89 | BenQ XL4230-B | 1920×1080 | 24 |
| Taiwan | National Taiwan Normal University | Mandarin | EyeLink 1000 | 20 | 60 | Dell P2417 | 1920 x 1080 | 24 |
| \*Turkey | Middle East Technical University | Turkish | Eye Link 1000+ | 20 | 60 | HP Pavilion 23cw | 1920 x 1080 | 23 |

**S3**: **Reliability estimates at the participant-level.** The table below presents reliability estimates at the participant-level for the eye-tracking data from the L2 reading task. To obtain these values, we calculated for each sample and for each eye-movement measure (except for reading rate) the correlation between mean values for 'odd' and 'even' words within a participant. Values before and after the comma represent uncorrected and Spearman-Brown corrected values, respectively. To obtain reliability estimates for reading rate, we used an Intra-class Correlation Coefficient (ICC) to examine the degree of agreement in reading rates across the 12 texts. Procedures are identical to those reported in Kuperman et al., 2023.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample L1** | *firstfix.dur* | *firstrun.dur* | *dur* | *nfix* | *reg.in* | *reread* | *skip* | *mean r*a | *Reading rate*b |
| Basque | 0.97, 0.98 | 0.96, 0.98 | 0.98, 0.99 | 0.97, 0.98 | 0.95, 0.97 | 0.98, 0.99 | 0.93, 0.96 | 0.97, 0.98 | 0.97 |
| Brazilian Portuguese | 0.97, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.97, 0.99 | 0.95, 0.98 | 0.97, 0.99 | 0.96, 0.98 | 0.97, 0.99 | 0.97 |
| Mandarin, simplified | 0.97, 0.99 | 0.97, 0.99 | 0.96, 0.98 | 0.97, 0.98 | 0.92, 0.96 | 0.97, 0.99 | 0.91, 0.95 | 0.96, 0.98 | 0.94 |
| Mandarin, traditional | 0.95, 0.97 | 0.98, 0.99 | 0.98, 0.99 | 0.97, 0.99 | 0.94, 0.97 | 0.98, 0.99 | 0.93, 0.97 | 0.97, 0.98 | 0.96 |
| Danish | 0.97, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.96, 0.98 | 0.99, 0.99 | 0.84, 0.91 | 0.97, 0.99 | 0.96 |
| English (UK) | 0.98, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.97, 0.99 | 0.97, 0.99 | 0.96, 0.98 | 0.93, 0.96 | 0.97, 0.99 | 0.97 |
| German (Potsdam) | 0.93, 0.96 | 0.93, 0.96 | 0.96, 0.98 | 0.96, 0.98 | 0.91, 0.95 | 0.96, 0.98 | 0.96, 0.98 | 0.95, 0.97 | 0.94 |
| German (Zurich) | 0.96, 0.98 | 0.98, 0.99 | 0.99, 0.99 | 0.98, 0.99 | 0.96, 0.98 | 0.97, 0.99 | 0.95, 0.97 | 0.97, 0.99 | 0.98 |
| Hindi (IIITH) | 0.99, 0.99 | 0.97, 0.99 | 0.98, 0.99 | 0.97, 0.98 | 0.94, 0.97 | 0.97, 0.99 | 0.95, 0.98 | 0.97, 0.99 | 0.95 |
| Hindi (IITK) | 0.97, 0.99 | 0.98, 0.99 | 0.99, 0.99 | 0.97, 0.99 | 0.91, 0.96 | 0.96, 0.98 | 0.95, 0.97 | 0.97, 0.98 | 0.97 |
| Icelandic | 0.98, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.95, 0.98 | 0.98, 0.99 | 0.94, 0.97 | 0.98, 0.99 | 0.97 |
| Norwegian | 0.96, 0.98 | 0.97, 0.98 | 0.98, 0.99 | 0.97, 0.99 | 0.96, 0.98 | 0.95, 0.97 | 0.93, 0.96 | 0.96, 0.98 | 0.98 |
| Russian | 0.99, 0.99 | 0.98, 0.99 | 0.99, 0.99 | 0.98, 0.99 | 0.94, 0.97 | 0.98, 0.99 | 0.95, 0.97 | 0.98, 0.99 | 0.97 |
| Serbian | 0.98, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.97, 0.98 | 0.96, 0.98 | 0.97, 0.98 | 0.91, 0.95 | 0.97, 0.98 | 0.97 |
| Spanish (Chile) | 0.98, 0.99 | 0.98, 0.99 | 0.99, ~1 | 0.98, 0.99 | 0.95, 0.97 | 0.97, 0.98 | 0.92, 0.96 | 0.97, 0.99 | 0.98 |
| Turkish | 0.99, 0.99 | 0.98, 0.99 | 0.98, 0.99 | 0.96, 0.98 | 0.97, 0.99 | 0.96, 0.98 | 0.87, 0.93 | 0.97, 0.98 | 0.96 |

Notes: firstfix.dur: first fixation duration; firstrun.dur: gaze duration; dur: total fixation time; nfix: number of fixations; reg.in: regression rate; reread: likelihood of second pass; skip: skipping rate; mean r: mean reliability across eye-tracking measures (excluding reading rate);

a Mean correlations are based on mean *z* values after Fisher r-to-z transformation, which were then transformed back to *r* values using an inverse transformation.

b Reading rate estimates are based on ICC. These values reflect the reliability of mean reading rate across the 12 texts at the participant-level.

**S4: Reliability estimates at the word token-level.** Below are the reliability estimates at the word token-level (i.e., the level of individual word occurrences) for the eye-tracking L2 reading task. These values were computed, for each language and eye-tracking measure, by examining the correlation between means for 'odd' and 'even' participants within each word token. Values before and after the comma represent uncorrected and Spearman-Brown corrected values, respectively. These procedures are identical to those reported in Kuperman et al., 2023.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample L1** | *firstfix.dur* | *firstrun.dur* | *dur* | *nfix* | *reg.in* | *reread* | *skip* | *mean r*a |
| Basque | 0.35, 0.52 | 0.61, 0.76 | 0.66, 0.79 | 0.66, 0.79 | 0.55, 0.71 | 0.45, 0.62 | 0.78, 0.88 | 0.59, 0.74 |
| Brazilian Portuguese | 0.37, 0.54 | 0.58, 0.74 | 0.63, 0.77 | 0.65, 0.79 | 0.64, 0.78 | 0.48, 0.65 | 0.84, 0.91 | 0.62, 0.76 |
| Mandarin, simplified | 0.39, 0.56 | 0.76, 0.86 | 0.75, 0.86 | 0.72, 0.84 | 0.50, 0.67 | 0.35, 0.51 | 0.85, 0.92 | 0.65, 0.78 |
| Mandarin, traditional | 0.34, 0.51 | 0.75, 0.86 | 0.78, 0.87 | 0.74, 0.85 | 0.52, 0.68 | 0.38, 0.55 | 0.79, 0.88 | 0.64, 0.78 |
| Danish | 0.31, 0.47 | 0.52, 0.68 | 0.55, 0.71 | 0.49, 0.66 | 0.50, 0.66 | 0.28, 0.43 | 0.73, 0.85 | 0.50, 0.66 |
| English (UK) | 0.42, 0.6 | 0.54, 0.70 | 0.55, 0.71 | 0.53, 0.69 | 0.65, 0.79 | 0.37, 0.54 | 0.86, 0.92 | 0.59, 0.74 |
| German (Potsdam) | 0.14, 0.25 | 0.33, 0.49 | 0.41, 0.58 | 0.48, 0.65 | 0.38, 0.55 | 0.24, 0.39 | 0.67, 0.80 | 0.39, 0.55 |
| German (Zurich) | 0.36, 0.53 | 0.64, 0.78 | 0.7, 0.82 | 0.70, 0.82 | 0.62, 0.76 | 0.46, 0.63 | 0.84, 0.91 | 0.64, 0.78 |
| Hindi (IIITH) | 0.42, 0.6 | 0.63, 0.77 | 0.68, 0.81 | 0.68, 0.81 | 0.64, 0.78 | 0.49, 0.66 | 0.88, 0.94 | 0.66, 0.79 |
| Hindi (IITK) | 0.48, 0.65 | 0.61, 0.76 | 0.67, 0.8 | 0.66, 0.80 | 0.56, 0.72 | 0.45, 0.62 | 0.87, 0.93 | 0.64, 0.78 |
| Icelandic | 0.47, 0.64 | 0.64, 0.78 | 0.66, 0.8 | 0.67, 0.80 | 0.63, 0.77 | 0.45, 0.62 | 0.85, 0.92 | 0.64, 0.78 |
| Norwegian | 0.16, 0.28 | 0.28, 0.44 | 0.37, 0.54 | 0.39, 0.57 | 0.36, 0.53 | 0.21, 0.34 | 0.57, 0.72 | 0.34, 0.50 |
| Russian | 0.33, 0.50 | 0.6, 0.75 | 0.65, 0.79 | 0.65, 0.79 | 0.56, 0.72 | 0.42, 0.59 | 0.83, 0.91 | 0.60, 0.75 |
| Serbian | 0.27, 0.43 | 0.51, 0.67 | 0.58, 0.74 | 0.65, 0.79 | 0.6, 0.75 | 0.38, 0.55 | 0.82, 0.9 | 0.57, 0.72 |
| Spanish (Chile) | 0.17, 0.3 | 0.49, 0.66 | 0.59, 0.74 | 0.63, 0.77 | 0.54, 0.7 | 0.39, 0.57 | 0.77, 0.87 | 0.53, 0.69 |
| Turkish | 0.2, 0.34 | 0.38, 0.55 | 0.38, 0.55 | 0.34, 0.51 | 0.34, 0.51 | 0.19, 0.32 | 0.64, 0.78 | 0.36, 0.53 |

Notes: firstfix.dur: first fixation duration; firstrun.dur: gaze duration; dur: total fixation time; nfix: number of fixations; reg.in: regression rate; reread: likelihood of second pass; skip: skipping rate; mean r: mean reliability across eye-tracking measures.

a Mean correlations are based on mean *z* values after Fisher r-to-z transformation, which were then transformed back to *r* values using an inverse transformation.

**S5: Reliability of other tests of individual differences**. The table below reports reliability estimates for measures from the battery of individual-differences tests in MECO L2 as well as for comprehension accuracy in the eye-tracking L2 reading task. For comprehension, spelling and motivation we calculated both split-half reliability and Cronbach's alpha. Only split-half reliability was calculated for the vocabulary task because of the task's adaptive nature (i.e., because different participants had data from different trials in this task as a function of when the stopping rule was applied). For the vocabulary task we provide reliability estimates for scores calculated according to performance in the full task, as well as for scores based on responses in 'thousands' 2-5 (see Kuperman et al., 2023, for details). For split-half reliability, uncorrected and corrected values appear before and after the comma, respectively.As expected given the use of identical tests and the large and generally similar samples, reliability estimates were comparable to those reported in Kuperman et al., (2023), with all values within +-0.1 in terms of (corrected) reliability estimates.

|  |  |  |
| --- | --- | --- |
| Measure | Split half | Cronbach's Alpha |
| Spelling | 0.60, 0.75 | 0.73 |
| Vocabulary: all word groups | 0.43, 0.60 | N/A |
| Vocabulary: word groups 2-5 | 0.58, 0.74 | N/A |
| Motivation | 0.51, 0.67 | 0.62 |
| Comprehension accuracy | 0.47, 0.64 | 0.61 |