**An Examination of professional translators in a real-world situation**

Matthias Kobi1, Michael Boos1, Stefan Elmer1\* & Lutz Jäncke1,2\*

1Division Neuropsychology, Department of Psychology, University of Zurich, Zurich, Switzerland.

2University Research Priority Program (URPP) “Dynamics of Healthy Aging”, University of Zurich, Zurich, Switzerland.

\* Shared last authorship

Keywords:

Generalized lexical decision task, word frequency, language switching, language expertise, diffusion modeling, EEG, N400

E-mail addresses:

MK: [matthias.kobi@psychologie.uzh.ch](mailto:matthias.kobi@psychologie.uzh.ch)

MB: m.boos@psychologie.uzh.ch

SE: s.elmer@psychologie.uzh.ch  
LJ: l.jaencke@psychologie.uzh.ch

Correspondence to:  
Matthias Kobi or Stefan Elmer  
Institute of Psychology  
Division Neuropsychology  
University of Zurich  
Binzmühlestrasse 14/25  
8050 Zurich, Switzerland

Date: 27 March, 2022

**Abstract**

**Introduction**

*Importance of bilingualism*

A significant part of about half of the world’s population in all age groups and socio-economic levels is multilingual and speaks two or more languages (Grosjean, 2010). In a globalized world, bilingualism and the need to communicate in a lingua franca are expected to become an even more widespread phenomenon. Additionally, an increasing part of the worldwide exchange is centered on written communication in several languages. Thus, in an academic, professional, or social environment, there is an increasing dependency on managing reading and writing skills in multiple languages. In this context, the physiological processes underlying visual word recognition as a critical reading component gained increased interest in several branches of research. In fact, while reading words or texts, physical signals (e.g., letters, symbols, and spaces) are used to build word-form representations that are subsequently mapped onto word candidates stored in long-term memory (Corona Dzul, 2017; Grainger & Jacobs, 1996). This process is called lexical access and involves identifying words in the mental lexicon based on structured sequences of letters (Dixon & Rothkopf, 1979).

Bilingualism can be used as a prime model to understand how multiple languages are represented and processed in the brain and how language proficiency influences word recognition (Abutalebi & Green, 2007; Consonni et al., 2013; DeLuca et al., 2019; Perani et al., 2003). Furthermore, the examination of bilingual individuals enables to differentiate between lexical access in L1 and L2 (Dijkstra & Kroll, 2005) and to assess symmetric or asymmetric processing costs while switching between the two languages (Abutalebi & Green, 2008; Declerck & Philipp, 2015; Luk et al., 2012). Currently, there is an ongoing discussion on whether lexical access in bilinguals is rather language-selective or language-nonselective (Dijkstra & Kroll, 2005). The hypothesis of selective lexical access postulates the existence of two or more independent lexica, and a linguistic stimulus only activates the corresponding language nodes. In contrast, the nonselective perspective acts on the assumption of integrated lexical access within a holistic bilingual lexicon, and each linguistic stimulus activates two or more languages simultaneously. The language-nonselective perspective is rooted in previous research on interlingual homographs, cognates, and interlingual neighbors, suggesting parallel lexical activation in both languages during bilingual word recognition (Dijkstra & Kroll, 2005). Furthermore, the nonselective nature of bilingual visual word recognition was implemented in a prominent localist-connectionist model called the “bilingual interactive activation” (BIA) model, which was extended later on to the BIA+ model (Dijkstra & van Heuven, 2002). In particular, the BIA+ model assumes two independent components, namely a word identification system and a task/decision system that acts as an executive control system. In a first step, the word identification system gathers information about a visual stimulus on different levels (i.e., sublexical and lexical orthography and phonology). Subsequently, possible word candidates activate the respective language affiliation nodes (i.e., language nodes) as well as the corresponding semantic representations. Accordingly, such a bottom-up activation of word representations enables parallel lexical activation and nonselective lexical access where word candidates from multiple languages with orthographic or phonological similarities, like for example cognates, can be activated during early processing stages. However, the same model also implies a top-down driven feedback mechanism where the activation of a language node (e.g., English) inhibits word candidates from other languages (Dijkstra & Kroll, 2005).

*Neurophysiological evidence of translation*

*How does expertise and experience affect the translating brain*

*Evidence from non-standard language input (ELF vs EdE)*

*Cognitive load*

*- Cognitive load theory*

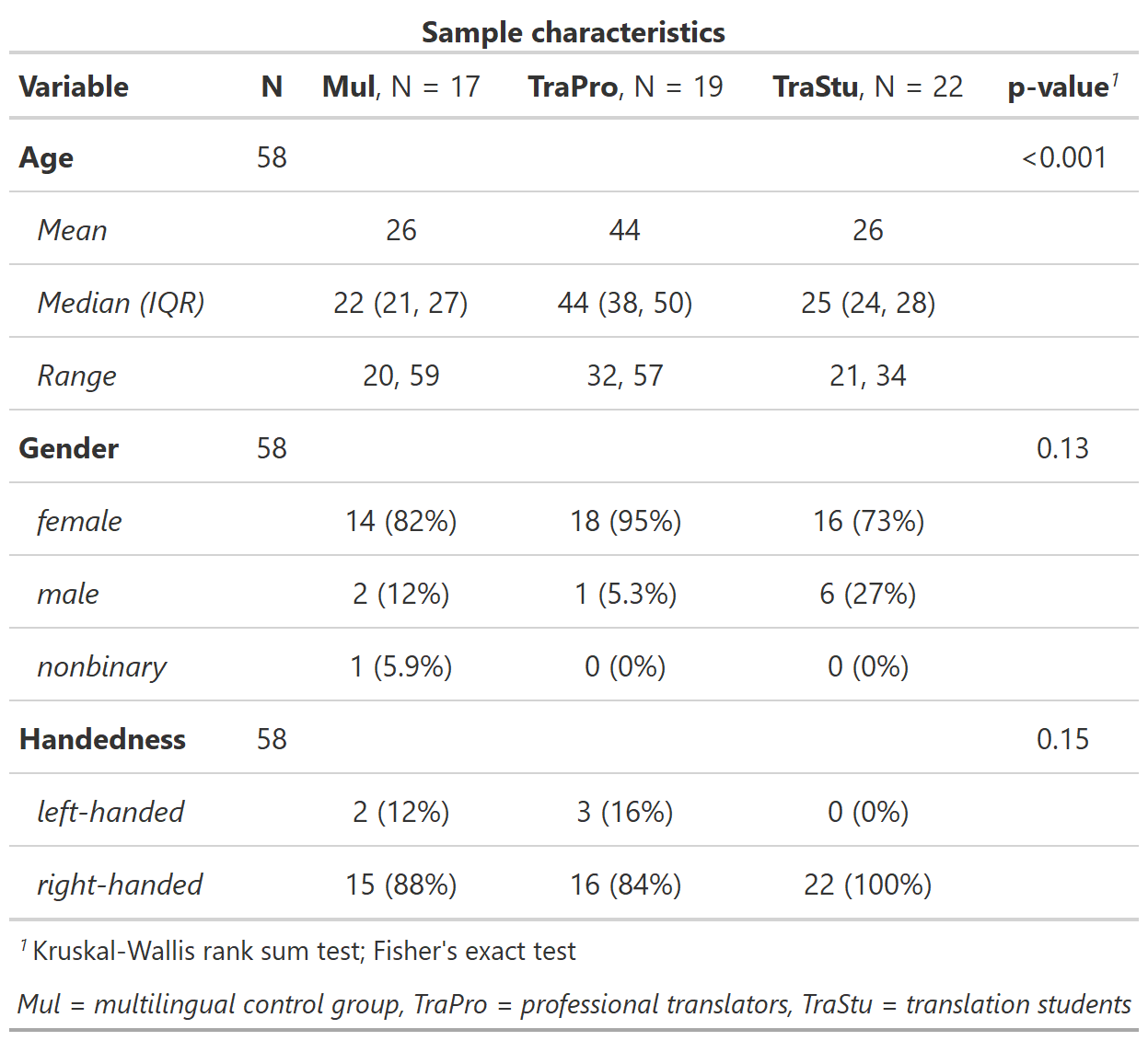
*- Measuring cognitive load with EEG*

*Hypothesis regarding our research*

**Methods**

**Participants**

We collected data from 72 native German (L1) participants (62 females, one nonbinary) with a professional English (L2) background. We recruited three different groups: professional translators (TraPro, N = 19), trainee translators (TraStu, N = 22), and a multilingual control group (Mul, N = 17). Members of the multilingual control group were students of English language and literature studies or teachers of English as a foreign language from high schools. All participants were required to use L2 in their daily routine, and primary direction of translating for professional and student translators was from L2 to L1. Since we recruited participants with varying levels of professional experience, groups could not be matched regarding their age (Table X) but they did not differ in respect of gender or handedness (Annett, 1970). All participants had normal or corrected-to-normal vision. Two participants reported using medicaments for diabetes, two for high blood pressure, and two participants reported concussions that occurred longer than five years before testing. The experiment lasted approximately four hours and was rewarded by cash. 14 participants were excluded from the analysis because of failure to follow the instructions of the experiment (N = 6), medication (N = 4, anti-depressants or Ritalin), and noisy or missing data (N = 4). Thus, we analyzed data from 58 participants. The study was carried out according to the principles in the declaration of Helsinki and approved by the Swiss National Science Foundation ethics committee.



*Table 1: Sample characteristics*

**Psychometrics and questionnaires**

Every participant completed a short English language test ([https://www.sprachtest.de/ einstufungstest-englisch](https://www.sprachtest.de/einstufungstest-englisch)) to provide an assessment of L2 proficiency. This online procedure lasted about 15 minutes and consisted of 13 vocabulary, grammar, listening, and reading comprehension questions. The maximum score of the test was 40. Furthermore, we collected data on the age of L2 acquisition and experience in translating and interpreting (cumulative training hours, and cumulative training hours per day since the age of 17) in a language background questionnaire. To assess working memory capacity, participants completed both a visual and an auditory 3-back task comprising of 60 letter stimuli, of which 20 were target stimuli. The order of the tasks was pseudorandomized across the groups. We analyzed N-back data using d-primes (d’). D-primes were calculated as the difference between the z-transformed hit rate and false alarm rate (Hautus et al., 2021). Additionally, we evaluated the cognitive capabilities using a short version of the WAIS (Wechsler Adult Intelligence Scale) test battery (Waldmann, 2008). This short version was composed of the following four subtests: number-symbol associations, detection of commonalities, the mosaic test, and digit span forward and backward. Using the standardized T-values, this composition sensitively reflects general intellectual abilities (Waldmann, 2008).

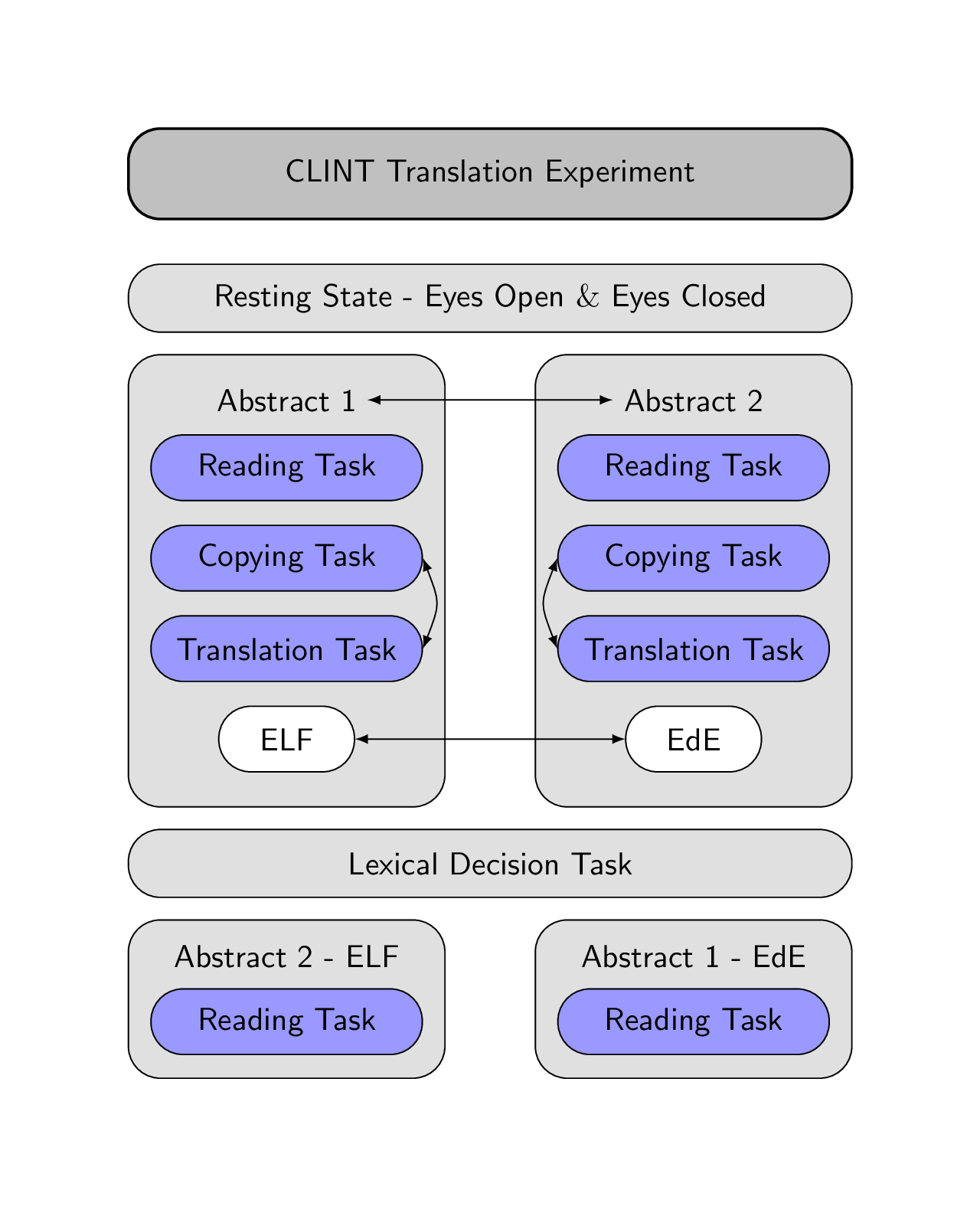
**Stimulus material**

In this experiment, we used two original abstracts in English that were submitted to conferences. Because the authors of the abstracts were non-native English writers, their original texts were regarded as ELF stimuli. In a next step, the two abstracts were processed into an edited-to-standard English (EdE) version by professional translators of the Zurich University of Applied Sciences (ZHAW). As little changes as possible were made to keep the text as close to the original while generating grammatically correct sentences and an overall better readability. This translation procedure resulted in four different text stimuli: text 1 (ELF, original), text 1 (EdE), text 2 (ELF, original), and text 2 (EdE).

*- how many sentences per version*

*- How many words per version*

**Experimental procedure**

****

*Figure 1: Experimental design. Arrows indicate randomizations in the task.*

After participants had completed the written informed consent, participants completed all psychometric measurements. First, participants completed an EEG resting state eyes open and eyes closed condition for three minutes each. Second, participants started with the reading task, followed by the copying and translation task for the first abstract. In the reading task, the text was presented sentence by sentence and participants were able to read at a self-paced speed through button press. Subsequently, we asked the participants how difficult they thought this task was and the answer was collected through mouse press on a 10 cm horizontal bar (easy on the left, difficult on the right). Additionally, we asked five control questions to check if participants have read the text. Those questions were multiple choice and participants had to choose one of three answer possibilities by pressing a key on the keyboard. In the copying task, participants were asked to copy the presented sentences and therefore, the generated output was English (L2). After completing a sentence, participants could move on to the next sentence by pressing “Enter”. In the translation task, the presented sentences were then translated to German (L1), reflecting a forward translation (FT). Again, after completion of a sentence, participants had to press “Enter” to continue with the next sentence. In all tasks, the words of the presented sentence were separated with double spacing and double lines. In the copying and translation task, the sentence that had to be processed was displayed in the upper part of the monitor while the answers of the participants were presented in the lower part. The duration of the reading task differed based on the self-paced reading of participants. However, the copying and translating task was limited to five minutes each. After working on the first abstract, participants continued using the same procedure with the second abstract.

In the experiment, we randomized the order of the abstracts (text 1, text 2), the version (ELF, EdE), and the copying and translation task across participants, indicated by the arrows in Figure X. Therefore. each participant processed an abstract only in one version but not in the other. If the first abstract was in ELF, the second was in EdE and vice versa. Since the duration of the copying and translation task was limited, participants did not process the whole text but always started from the beginning and worked through the text sentence by sentence. However, it was made sure that no sentence was used twice in the copying and translation task.

After processing the two abstracts, participants completed a lexical decision task and then had to read the abstracts again in the other versions. However, we did not include data from those two conditions in the analyses. To start the experiment, instructions to the task were presented on the computer screen, and to become confident with keyboard, participants had to copy a sentence that contained all possible special symbols from the abstract.

**Behavioral data processing**

The preprocessing of the behavioral data was done using R (version 3.6.3, <https://www.r-project.org/>). For the reading task, we evaluated the percentage of the correct answered control questions per text as well as the average reading duration per sentence, which was adjusted for the different lengths of the texts. Furthermore, we analyzed the perceived difficulty as the distance in cm of the mouse click from 0 (easy) for both the reading and translation task. Therefore, higher values indicate a higher perceived difficulty of the task. Regarding the copying and translation tasks, we evaluated the total amount of chars typed during the 5 minutes, as well as the number of chars typed if deletions were subtracted (chars endversion). Those variables can be regarded as a measure for the efficiency in the copying and translation task. Furthermore, we retrieved the number of deletions for both tasks which refer to the total number pressing the “backslash” on the keyboard. Finally, we analyzed the output generated by each participant in the translation task by rating the fluency (0: incomprehensible – 5: flawless German) and the accuracy ( 0: no meaning – 5: all meaning) per sentence (Koehn & Monz, 2006). For the rating, we fully randomized the sentences from both texts and conditions as well as all participants. Three independent raters (*language experts from the from the IUED Institute of Translation and Interpreting of the ZHAW*) first rated the fluency and subsequently the accuracy of all sentences. For the accuracy rating, the translation output was compared to a reference translation provided by the IUED. Then, the ratings were averaged per condition (ELF vs. EdE) and both texts to calculate an intraclass correlation coefficient (ICC) using the irr package (Version 0.84.1, <https://cran.r-project.org/web/packages/irr/>) in R. Applying a 2-way mixed-effects model of the type “consistency” and a mean-rating (k=3) revealed a ICC(C,3) = 0.575 (95%-confidence interval = 0.421 – 0.694) for the fluency rating and a ICC(C,3) = 0.909 (95%-confidence interval = 0.875 – 0.934) for the accuracy rating. The ICC for the fluency rating likely reflects a moderate reliability, whereas the ICC for the accuracy rating reflects an excellent reliability (Koo & Li, 2016). Finally, we averaged the three raters to generate a score for a mean rater further used in the statistical analyses.

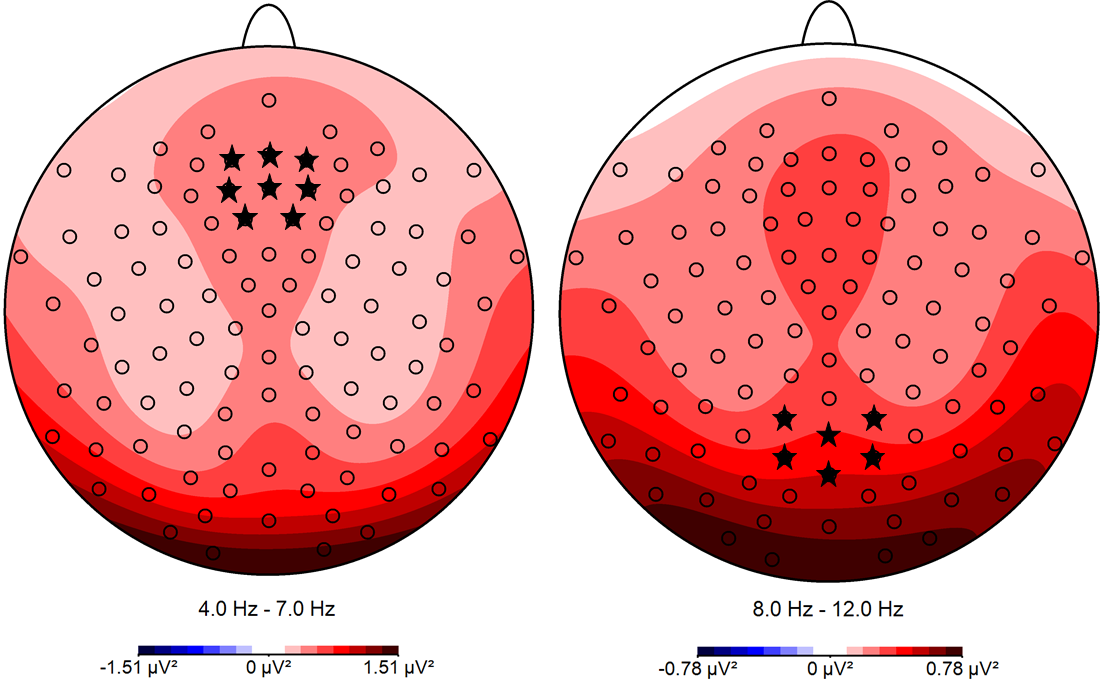
**EEG data acquisition**

EEG data acquisition took place in a light-dimmed Faraday cage where the participants were seated approximately 70 cm in front of a 24-inch monitor. The participants were instructed to relax and stay as still as possible during the EEG measurements. The experiment was programmed in MATLAB 2016b using the Psychophysics Toolbox Version 3 extension (Kleiner et al., 2007) for behavioral data acquisition. We recorded high-density EEG data at a sampling rate of 500 Hz with a bandpass filter of 0.1-100 Hz using the EGI 300 Geodesic EEG system with a 128-channel HydroCel Geodesic Sensor Net (HCGSN) (Electrical Geodesics, Eugene, Oregon). Before recording, each electrode was double-checked to ensure good contact on the scalp, and impedances were kept below 40 kOhm. This procedure was repeated after the EEG resting state, the processing of abstract 1 and 2, and the lexical decision task. The recording reference electrode was Cz.

**EEG data processing**

The data was processed using MATLAB2018b, EEGLAB 2021\_0, and Brain Vision Analyzer version 2.2.0 (BrainProducts, Munich, Germany). For EEG data preprocessing, we used the Automagic toolbox implemented in MATLAB (v.2.5, Pedroni, Bahreini, & Langer, 2019), which is a pipeline for automatic EEG data cleaning. First, the number of EEG channels was reduced to 105 by discarding channels lying on the neck and face. Second, we applied the PREP pipeline for bad channel detection with the minimum variance set to 1. Third, we used the ICLabel approach with a temporary 2 Hz high-pass filter for artifact correction to remove muscle, eye, heart, and channel noise components with a probability threshold higher than 0.8. Fourth, we selected eleven frontal electrodes for the electrooculogram (EOG) regression. Fifth, power line noise was removed using the ZapLine method (de Cheveigné, 2020), eliminating five components. Sixth, we applied a 0.1 Hz high-pass and a 30 Hz low-pass filter. Finally, bad channels were reconstructed through spherical interpolation, and we applied the detrending algorithm to remove slow drifts. This procedure was applied independently for each abstract to avoid preprocessing unnecessary noise during pauses.

Further preprocessing of the EEG data was executed in the Brain Vision Analyser. First, we segmented the EEG into the different task segments and re-referenced the data to an average reference montage. Second, we used an automatic raw data inspection to mark bad time windows indicating remaining artifacts that were not removed by Automagic. Third, we divided the data into segments of 2 s length without overlaps, in which data segments marked as bad were skipped. Forth, a fast fourier transform (FFT) with a Hanning window (Length = 10%) was applied on all remaining segments. The resulting transforms were averaged per participant and per condition before exporting the power values for the theta (bandwidth from 4 to 7 Hz) and alpha (bandwidth from 8 to 12 Hz) band. Based on the voltage distribution of the grand average across the reading, copying and translation tasks, we analyzed theta power at a frontal (E4, E5, E10, E11, E12, E16, E18, and E19) and alpha power at a parietal (E61, E62, E67, E72, E77, and E78) electrode pool (see Figure X for electrode positions). Thus, we averaged the power per pool and frequency band for statistical analysis.

**

*Figure X: Topographical voltage distribution maps for theta and alpha band across reading, copying and translation tasks, and all participants. The channels selected for analyses are marked with \*.*

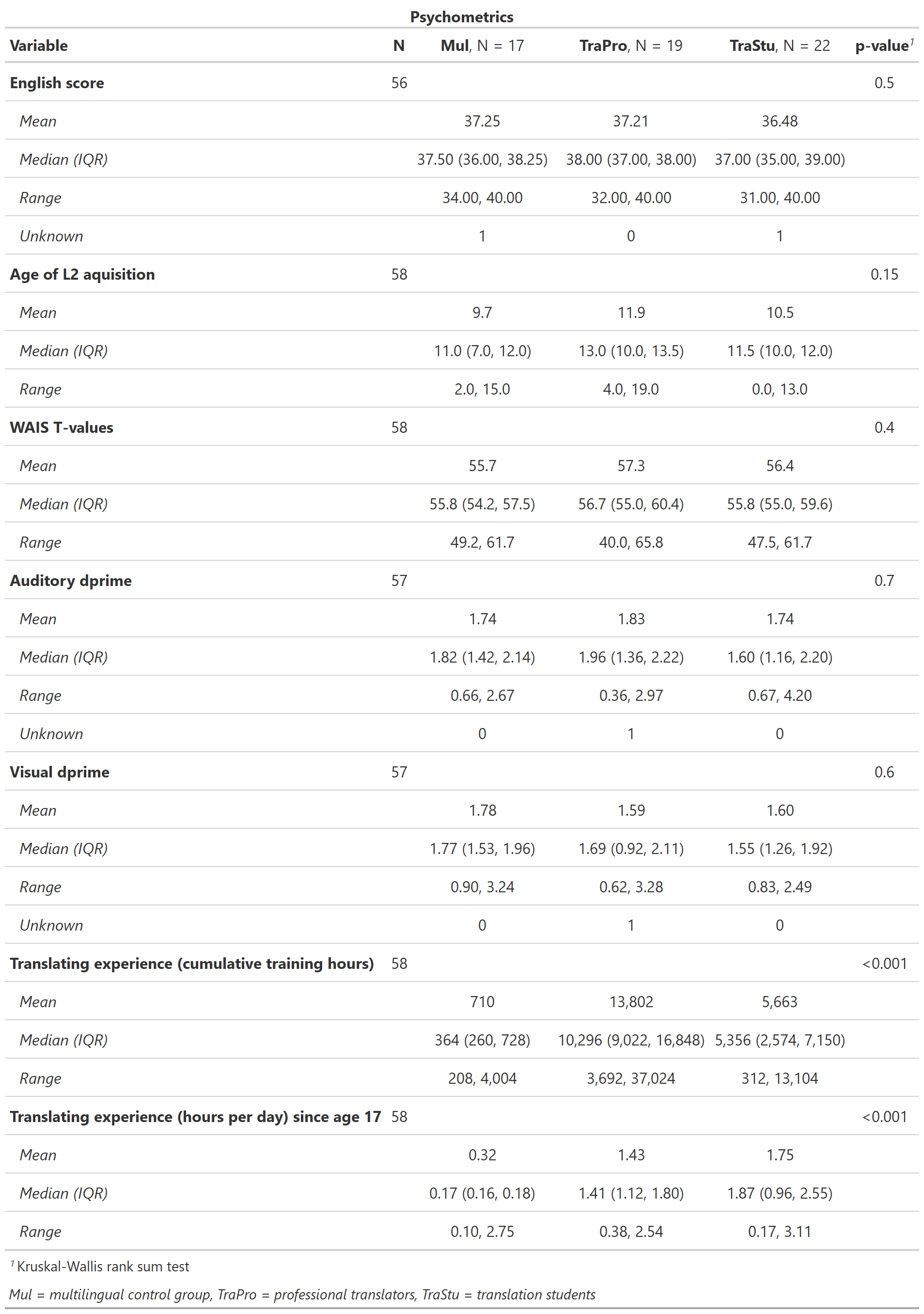
**Statistical analyses**

All statistical analyses were performed using Linear Mixed Models (LMM) implemented in the lme4 package (Version 1.1-23, <https://cran.r-project.org/web/packages/lme4/>) in R. For model-fitting, we used a bottom-up strategy starting with the null model and added random and fixed effects for our target variables. In general, we used three levels group (TraPro, TraStu, and Mul), three levels for task (reading, copying, and translation task), two levels for text (text1 and text2) as well as two levels for condition (EdE and ELF).

**Results**

**Psychometrics and questionnaires.**

Our groups did not differ regarding English score, age of L2 acquisition, WAIS T-values, auditory d’, and visual d’ (Table X). However, as expected by our recruitment, our groups differed regarding cumulative training hours in translating and interpreting (TraPro: M = 13’802, TraStu: M = 5’663, Mul: M = 710), as well as in the cumulative training hours per day since the age of 17 (TraPro: M = 1.43, TraStu: M = 1.75, Mul: M = 0.32).

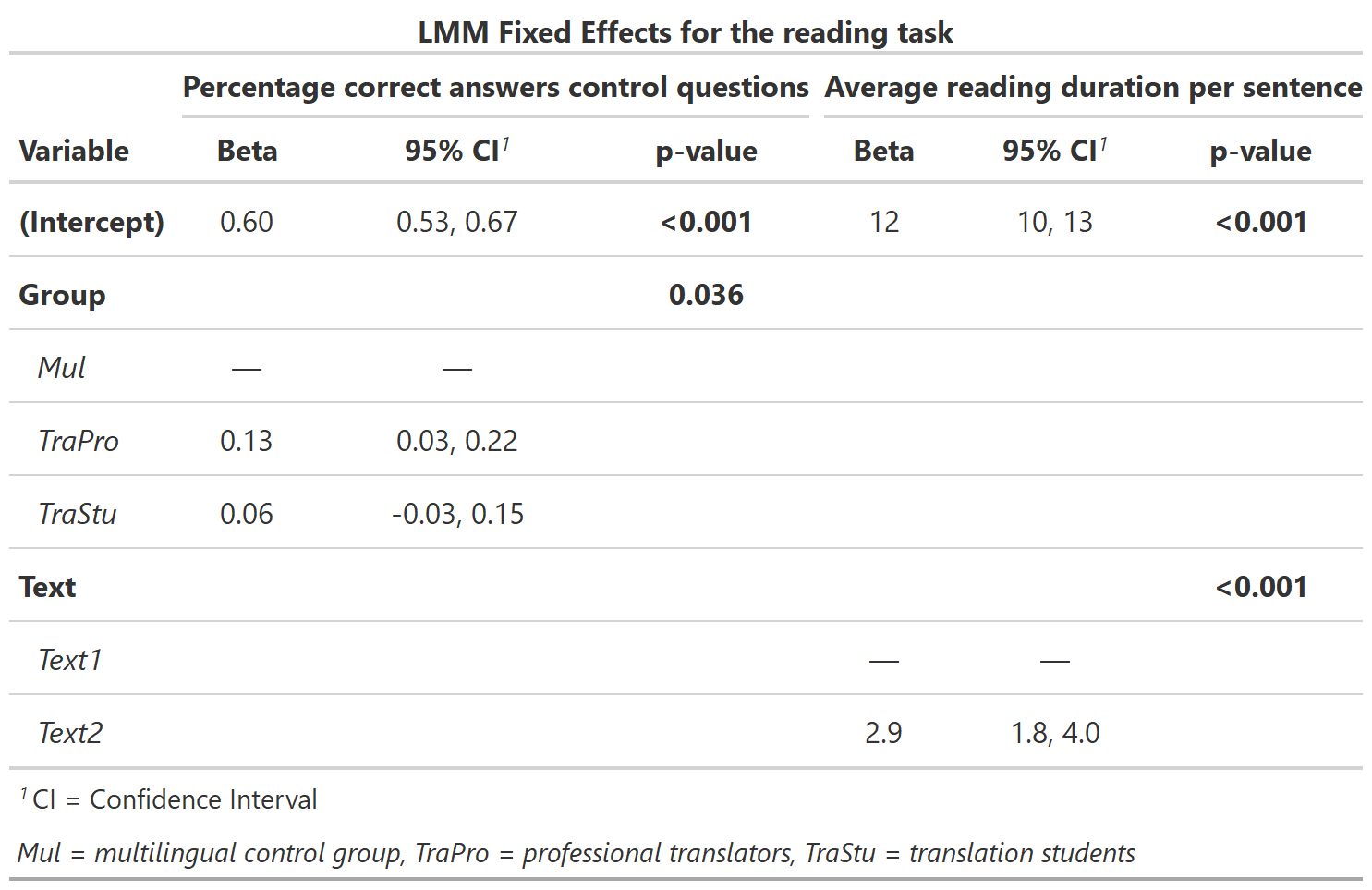
****

*Table X: Results of the psychometrics and questionnaires.*

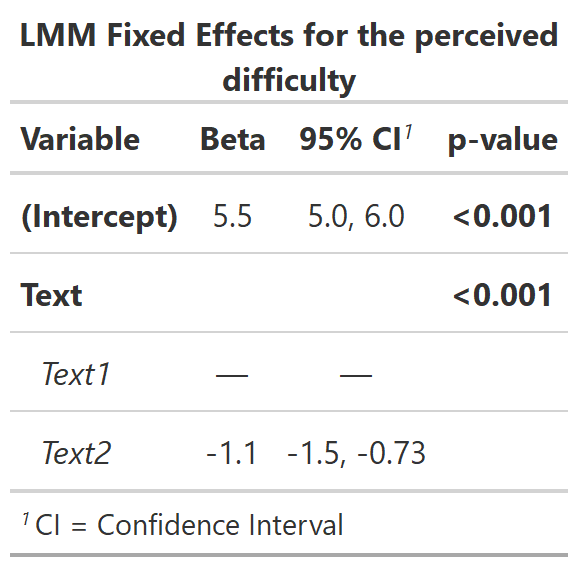
**Behavioral results**

**Reading task**

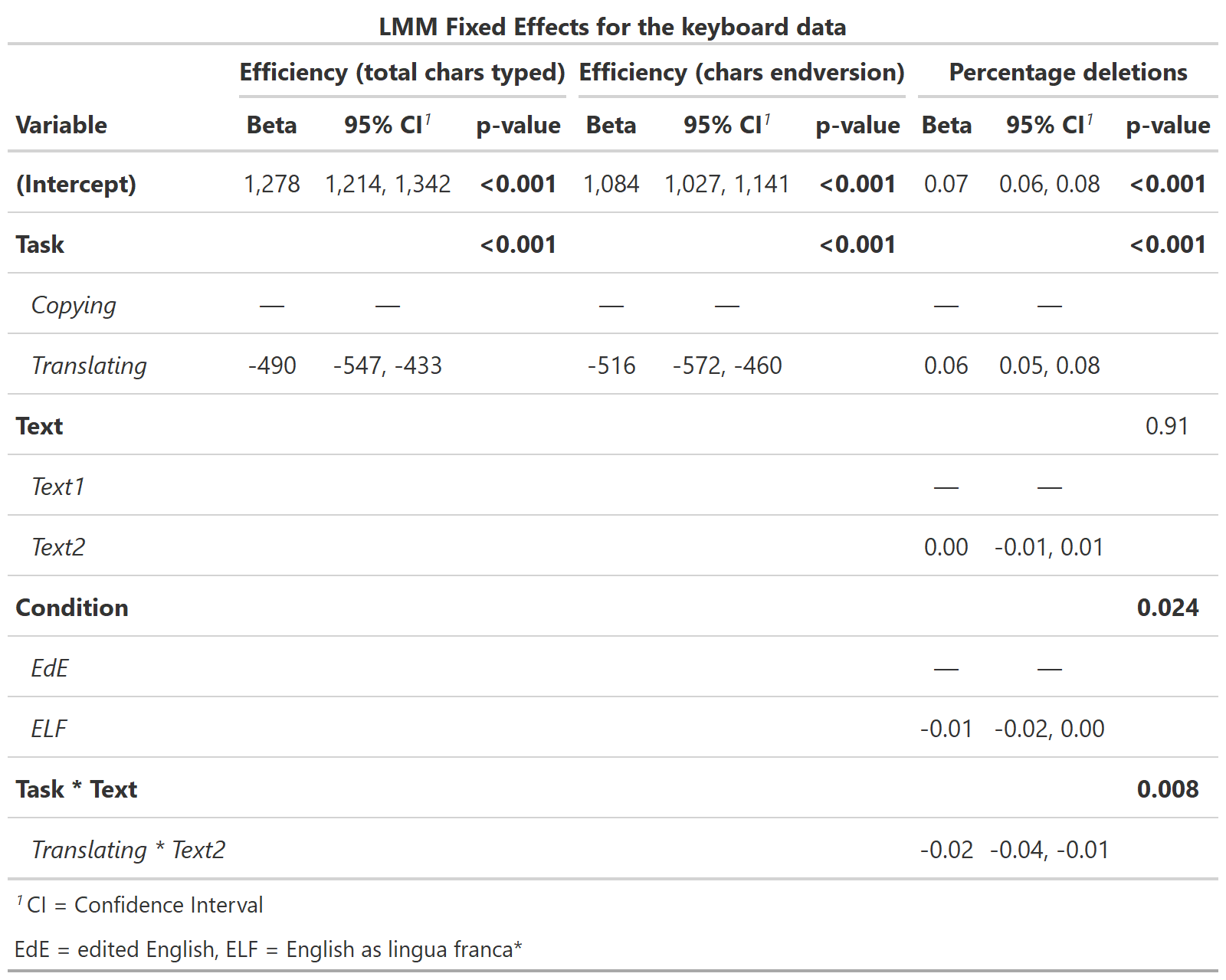
One limitation of our study is that we could only include a relatively low number of trials per

****

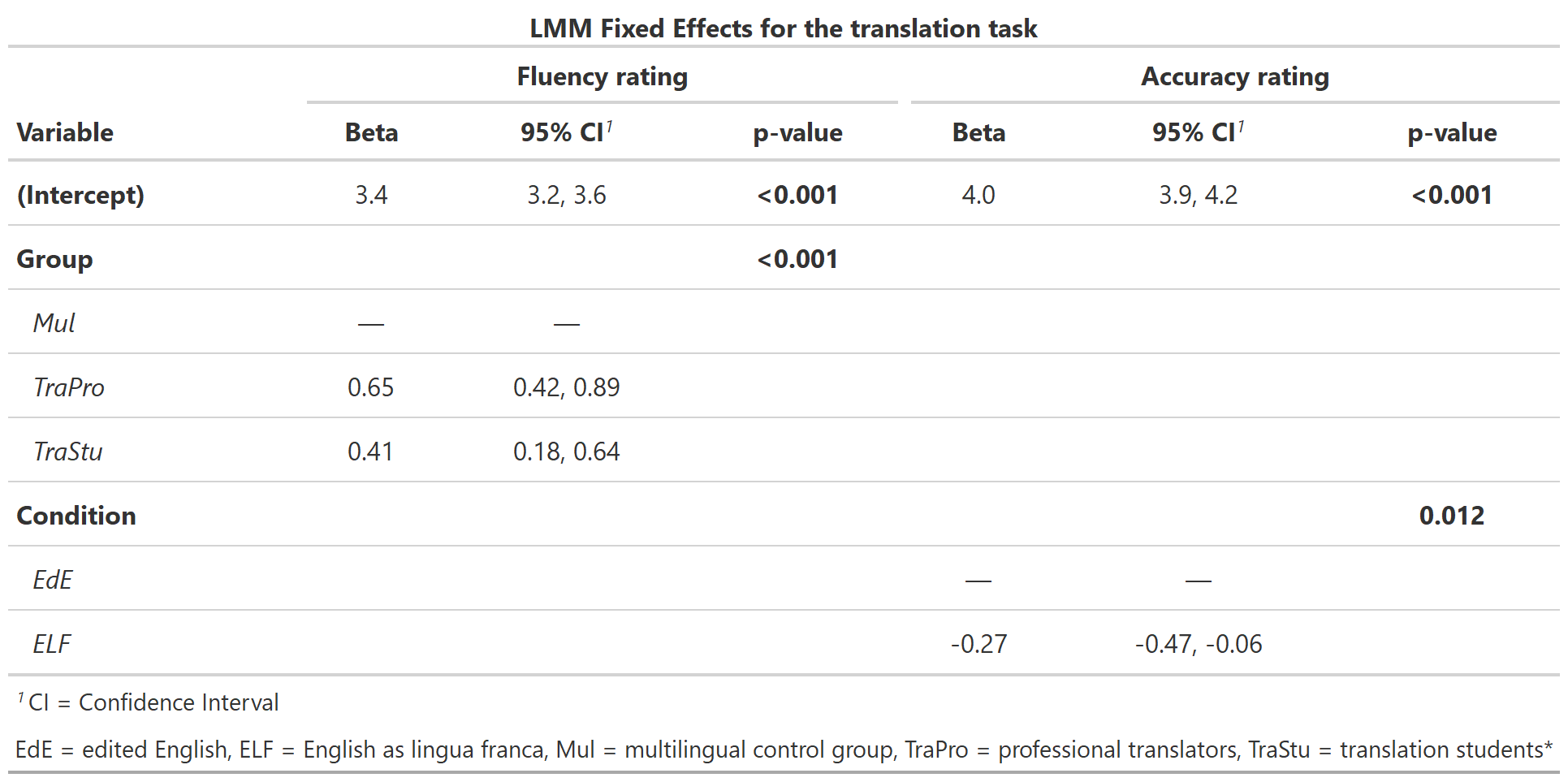
**Perceived difficulty of the reading and translation task**

****

**Keyboard data**

****

**Translation task**

****

**Discussion**

**Limitations:**

One limitation of our study is that we could only include a relatively low number of trials per condition, namely 60 in the word frequency tasks and 54 in the switching task. After having excluded trials with a high variance for the ERPanalyses, some participants with less than 40 trials may have resulted in a lower precision and stability for the extraction of the N400 features with the Liesefeld approach. Furthermore, using the Maximum Likelihood approach to estimate DM parameters could have led to less precise results. A further limitation of our word frequency task is that we did not control for factors that are known to have an influence on the word frequency effect, such as age when a word was learned, similarity to other words, word prevalence, contextuality, vocabulary size, and word knowledge (Brysbaert et al., 2018; Juhasz et al., 2018). Another limitation is that we recruited German-English bilinguals and native German participants who learned English as a second language. For the latter, information about participants’ age of acquisition and exposure to English was not collected. Furthermore, English proficiency was tested using a short screening of language competence which could have resulted in a ceiling effect, especially for high-scoring participants. Finally, it is important to mention that the factor of age was not normally distributed in our sample because we included more younger participants.

**Conclusions**

**References**

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242–275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, *23*(4), 557–582. https://doi.org/10.1080/01690960801920602

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, *61*(3), 303–321. https://doi.org/10.1111/j.2044-8295.1970.tb01248.x

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Consonni, M., Cafiero, R., Marin, D., Tettamanti, M., Iadanza, A., Fabbro, F., & Perani, D. (2013). Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex*, *49*(5), 1252–1258. https://doi.org/10.1016/j.cortex.2012.04.009

Corona Dzul, B. (2017). *Visual word recognition in bilinguals and monolinguals: behavioural and ERP investigations of the role of word frequency, lexicality and repetition*. University of Nottingham.

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, *207*. https://doi.org/10.1016/j.neuroimage.2019.116356

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin and Review*, *22*(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7565–7574. https://doi.org/10.1073/pnas.1811513116

Dijkstra, T., & Kroll, J. F. (2005). Bilingual visual word recognition and lexical access. In *Handbook of bilingualism: Psycholinguistic approaches* (178th ed., p. 201).

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/s1366728902003012

Dixon, P., & Rothkopf, E. Z. (1979). Word repetition, lexical access, and the process of searching words and sentences. *Journal of Verbal Learning and Verbal Behavior*, *18*(5), 629–644. https://doi.org/10.1016/S0022-5371(79)90354-2

Grainger, J., & Jacobs, A. M. (1996). Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model. *Psychological Review*, *103*(3), 518–565. https://doi.org/10.1037/0033-295X.103.3.518

Grosjean, F. (2010). Bilingual: Life and reality. In *Bilingual: Life and reality.* Harvard University Press. https://doi.org/10.4159/9780674056459

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). Detection theory: A user’s guide. In *Routledge*. https://doi.org/10.4324/9781003203636

Juhasz, B. J., Yap, M. J., Raoul, A., & Kaye, M. (2018). A Further Examination of Word Frequency and Age-of-Acquisition Effects in English Lexical Decision Task Performance: The Role of Frequency Trajectory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 85. https://doi.org/http://dx.doi.org/10.1037/xlm0000564 This

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & Cornelissen, F. (2007). *What’s new in Psychtoolbox-3?* https://pure.mpg.de/rest/items/item\_1790332/component/file\_3136265/content

Koehn, P., & Monz, C. (2006). *Manual and Automatic Evaluation of Machine Translation between European Languages 1 Evaluation Framework*. 102–121. http://www.statmt.org/wmt06/

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/J.JCM.2016.02.012

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, *27*(10), 1479–1488. https://doi.org/10.1080/01690965.2011.613209

Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., & Fazio, F. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, *19*(3), 170–182. https://doi.org/10.1002/hbm.10110