New Data on Text Reading in English as a Second Language: The Wave 2 Expansion of the Multilingual Eye Movements Corpus (MECO)

Victor Kuperman[1(*)], Sascha Schroeder[2(*)],

Cengiz Acartürk[3], Niket Agrawal[4], Lena S. Bolliger[5], Jan Brasser[5], César Campos-Rojas[6,7],

Denis Drieghe[8], Dušica Filipović Đurđević[9,10], Sofya Goldina[11,12], Romualdo Ibáñez Orellana[6,7],

Lena A. Jäger[5,13], Ómar I. Jóhannesson[14], Anurag Khare[4], Nik Kharlamov[15], Hanne B. S. Knudsen[15],

Árni Kristjánsson[14], Charlotte E. Lee[8], Jun Ren Lee[16], Marina P. T. Leite[17], Simona Mancini[18,19],

Nataša Mihajlović[10], Ksenija Mišić[9], Miloslava Orekhova[12], Olga Parshina[20,12],

Milica Popović Stijačić[21,10], Athanassios Protopapas[22], David R. Reich[13], Anurag Rimzhim[23],

Rui Rothe-Neves[17], Thais M. M. Sá[24], Andrea Santana Covarrubias[6], Irina Sekerina[25],

Heiða M. Sigurðardóttir[14], Anna Smirnova[26,12], Priyanka Srivastava[27], Ivana Ugrinic[22], Kerem Alp Usal[28],

Karolina Vakulya[29], Ark Verma[4], Denise H. Wu[30], Jin Xue[31,32], Sunčica Zdravković[10,9], Junjing Zhuo[33,32],

Laoura Ziaka[22,34], and Noam Siegelman[35(*)]

[1]McMaster University [2]University of Goettingen [3]Jagiellonian University [4]Indian Institute of Technology Kanpur [5]University of Zurich [6]Pontificia Universidad Católica de Valparaíso [7]Millennium Nucleus for the Science of Learning (MiNSoL) [8]University of Southampton [9]University of Belgrade [10]University of Novi Sad

[11]Université Paris Cité [12]Higher School of Economics (HSE) University Moscow [13]University of Potsdam [14]University of Iceland [15]Aalborg University [16]National Taiwan Normal University [17]Universidade Federal de Minas Gerais [18]Basque Center on Cognition Brain and Language [19]Ikerbasque, Basque Foundation for Science [20]Middlebury College [21]Singidunum University [22]University of Oslo [23]College of the Holy Cross [24]Universidade Federal de Lavras, [25]College of Staten Island of the City University of New York [26]University of Groningen [27]International Institute of Information Technology Hyderabad [28]Middle East Technical University [29]University of Plymouth [30]National Central University [31]Beijing Institute of Technology [32]University of Science and Technology Beijing [33]Northeast Normal University [34]Oslo University Hospital [35]Hebrew University of Jerusalem

Corresponding Author:

Victor Kuperman

Togo Salmon Hall 510
Department of Linguistics and Language, McMaster University
1280 Main Street West
Hamilton, Ontario, Canada, L8S 4M2

E-mail: vickup@mcmaster.ca

(*) V. Kuperman, S. Schroeder, and N. Siegelman contributed equally to this work.

New Data on Text Reading in English as a Second Language: The Wave 2 Expansion of the Multilingual Eye Movements Corpus (MECO)

Abstract

This paper reports an expansion of the English as a Second Language (L2) component of the Multilingual Eye Movement Corpus (MECO L2), an international database of eye movements during text reading. While the previous Wave 1 of the MECO project (Kuperman et al., 2023) contained English L2 reading data from readers with 12 different L1 backgrounds, the newly collected dataset adds eye-tracking data on English text reading from 13 distinct first-language (L1) backgrounds (N = 661), as well as participants' scores on component skills of English proficiency and information about their demographics and language background and use. The paper reports reliability estimates, descriptive statistics, and correlational analyses, as means to validate the expansion dataset. Consistent with prior literature and MECO's Wave 1, trends in the MECO Wave 2 data include a weak correlation between reading comprehension and oculomotor measures of reading fluency, and a greater L1-L2 contrast in reading fluency than reading comprehension. Jointly with Wave 1, the MECO project includes English reading data from more than 1200 readers representing a diversity of native writing systems (logographic, abjad, abugida, alphabetic) and 19 distinct L1 backgrounds. We provide multiple pointers to new venues of how L2 reading researchers can mine this rich publicly available dataset.

Keywords: Reading; second language; eye-tracking; comprehension; fluency.

While highly prolific, research into bilingualism and second language (L2) reading represents a relatively small subset of first (L1) and additional languages (for estimates see e.g., Melby-Lervåg & Lervåg, 2014; Siegelman et al., 2023). Arguably, the need for a broader coverage is particularly felt in the research stream that uses eye-tracking to study second language reading behavior. Because of the relatively high cost of eye-tracking equipment, this type of experimentation is largely concentrated in high-income countries with developed scientific infrastructure (e.g., Godfroid, 2020). Thus, existing eye-tracking studies in this field are biased towards L2s that are official languages of so-called WEIRD (i.e., Western, Educated, Industrialized, Rich, and Democratic) societies and the L1s that are well represented among international university students or immigrants in WEIRD countries. More broadly, L2 reading research is in constant need of methodologically comparable, high-quality, empirical data (see discussions in De Bruin, 2019; Gullifer & Titone, 2020; Luk & Bialystok, 2013, among many others), and again this is arguably particularly true in research into eye-movements in L2 given the still limited coverage of this line of research.

One recent approach to addressing these needs has emerged in the form of mega-studies that coordinate data collection across multiple labs worldwide, using comparable texts, reader populations, and procedures (see Brysbaert & Drieghe, 2024, for discussion). One such study (Kuperman et al., 2023) presents eye-tracking data on English text reading produced by N=543 L1 speakers of 12 languages (for other examples, see Berzak et al., 2022; Cop et al., 2017; Siegelman et al., 2023; Sui et al., 2023), along with several tests of component skills of English proficiency and rich demographic and language-background data. This study is one component of the Multilingual Eye Movements Corpus (MECO), labeled MECO L2. Within the MECO project, the same participants (at a given data collection wave) produced eye-tracking data on text reading in their L1 (MECO L1, Siegelman et al., 2022) and in English, which enables within-participant comparisons of oculomotor behavior in one's L1 and L2.

While the published Wave 1 of the MECO project in Kuperman et al. (2023) has already provided a solid expansion of the empirical base for studies of L1 and L2 reading, the current paper makes a further contribution to existing research needs, reporting new eye-tracking and skill test data on L2 reading that constitutes Wave 2 of the MECO L2 project. The first major goal of the current paper is to expand the coverage of the MECO L2 project in terms of the language background represented in the database. Thus, here we report data from 16 samples, representing 13 distinct L1 backgrounds, contributing eye-tracking data on L2 English text reading, along with measures of English component skills and language and demographic background. Most samples in the current wave are from sites where participants' L1 background is new to the MECO project: Specifically, we add a total of nine new samples of participants with seven L1 backgrounds previously uncovered in the MECO Wave 1: i.e., Basque, Brazilian Portuguese, Danish, Hindi, Icelandic, Mandarin (both simplified and traditional script), and Serbian. Each of these sites aimed to include a minimum of N=45 usable participants, and in most cases, samples met this threshold (see below). As a result, taken jointly, Waves 1 and 2 of MECO L2 bring the total of different L1 backgrounds in the English reading data from 12 (reported in Kuperman et al., 2023) to 19. Given that all measures and procedures reported here are fully comparable with those in the previous Wave of the project (Kuperman et al., 2023), the full database now presents researchers with an unprecedented opportunity to examine the determinants of English L2 proficiency and fluency across a very wide range of participant language backgrounds. The scope of the database enables tackling many novel theoretical questions, including, for example, questions about the links between eye-movement behavior and component skills of English reading, and the language distance between the L1 background of the reader and English as L2. Another benefit of the Wave 2 expansion is an addition of native readers of very different writing systems from the alphabetic system of English, i.e., Chinese and Hindi (in addition to the Korean Hangul and Hebrew abjad represented in Wave 1). This increased diversity of writing systems enables users of the MECO database to systematically study the

effects of the writing system on English reading proficiency as well (e.g., Bialystok et al., 2005; Geva & Siegel, 2000). Clearly, the questions mentioned here are simply examples of the types of investigations made possible with the full MECO L2 data. The major goal of this paper is to present and validate this rich dataset and make it publicly available for researchers for secondary use in line with open science practices and mega-studies of reading and language.

A second crucial goal of Wave 2 of the MECO L2 project is to increase the sample size of the MECO L2 database, in order to improve statistical power of studies using the open dataset. As discussed in detail in Kuperman et al., 2023, the MECO database is structured to enable both "bird's-eye view" type of analyses of similarities and differences in reading behavior across many language backgrounds, as well as targeted analyses of data from specific sites, theoretically interesting L1 pairs/groups, or specific L1 families (see also Siegelman et al., 2022, for a related discussion in the context of the MECO L1 component). With the new addition of the current Wave 2 data to the MECO L2 component, researchers will now have access to an unprecedented number of N=1204 participants reading in English across the project's two waves (i.e., adding N=661 new participants to the N=543 in Kuperman et al., 2023). This addition, combined with the improved cross-linguistic coverage discussed above, may also substantially improve analyses targeting specific L1 groups or typological families. In this context, note that the current Wave 2 data includes additional data for two samples included in the Wave 1 project, the Turkish and Norwegian samples, where data collection of Wave 1 was interrupted by the closures related to the COVID-19 pandemic. Below we refer to these two samples as "appended samples". The addition of new participants to these two samples was meant to make sure their sample size is in line with the target number of participants per MECO site (N=45 or more, see below).

Finally, another goal of the current Wave 2 data is to make available several "replication samples", i.e., data records that represent L1 backgrounds already found in Wave 1 but collected at different universities or countries. Thus, the Waves 1 and 2 of MECO L2 jointly include three samples of

participants with German as L1 (two from Germany and one from Switzerland), two with English (from Canada and UK), two with Hindi (both from India), two with Russian (both from Russia), and two with Spanish as L1 (from Argentina and Chile). The replication samples enable methodologically important comparative analyses of multiple samples from the same language background. Such analyses make it possible to disentangle the effect of the language background and the effect of the specific sample, university, or country. Also, they can be used to determine whether readers with the same L1 background are more similar to one another in their English L2 reading proficiency than speakers with different L1 backgrounds.

With these goals in mind, in the current paper we present the MECO L2 Wave 2 data. We start by providing full information about the included participants, eye-tracking methodology and procedure, tests of component skills, and questionnaire data collected. We then follow with analyses of the reliability of the collected data, as well as descriptive information regarding the distribution of basic eye-movement measures and measures of component skills across sites. These are meant to establish the collected database as a useful tool that can form the basis for secondary analyses in future research. We end with a few pointers to the future directions that L2 reading research can take with the help of the newly expanded MECO database.

Method

Participants. The present data on reading in English – labeled Wave 2 of the MECO L2 database – stem from 16 eye-tracking university-based laboratories in Asia, Europe, and South America. English was the first and dominant language for only one of the partner sites (UK), while the first and dominant languages in other samples was the official language of university instruction (and typically, the official

language of the country)[1]. All participants were university students or (rarely) staff members. With the present emphasis on the typical English as L2 readers, we applied a screening procedure that also took place in Wave 1 of the project (Kuperman et al., 2023). Specifically, we excluded participants with uncharacteristically high English fluency in all but the UK-based sample, i.e., self-reported simultaneously bilingual participants (with English as one of the languages), majors in English language or literature, and individuals who have lived for more than six months in an English-speaking country. The ethics clearance was obtained by each participating site from the ethics research board of the corresponding institution or country. Complete details of participant recruitment, materials, procedure, and apparatus of the present study are highly compatible with those used during Wave 1 of MECO data collection, see Kuperman et al. (2023): Our description below draws relevant details from Kuperman et al.'s Method section.

Table 1 lists the country and institution where the data were collected, sample size, and details regarding the participants' compensation, as well as the L1, age, and years of education of participants. Complete demographic information can be found in the project's data repository (see *Data availability*). In total, the current Wave 2 of MECO includes 661 new participants with valid eye-tracking data.

[1] Although the first and dominant language was Hindi for the Indian samples, English was their official language of university-level instruction. Further, in India, many schools and higher educational institutes teach in English, therefore, most participants had already received education in English from primary-school level onwards. Also note that in the Basque country, Spanish and Basque are both official languages of university instruction.

Table 1. Information regarding participants in available samples.

Country	Sample Code	Institute	L1	N: L2 data	Mean Age (range)	Mean Years of Education (SD)	Participants' compensation	Texts after trimming, %	Word tokens after trimming
Brazil	bp	Federal Universities of Ceara and Minas Gerais	Portuguese	54	22.00 (18-30)	17.65 (3.37)	Volunteer	69	62608
China	ch_s	University of Science and Technology Beijing	Mandarin (simplified script)	47	22.98 (20-30)	16.83 (2.37)	70 RMB / hour	64	48017
Chile	sp_ch	Pontificia Universidad Católica de Valparaíso	Spanish	45	21.71 (18-31)	15.27 (2.19)	Volunteer	66	49434
Denmark	da	Aalborg University	Danish	25	23.10 (19-30)	14.54 (1.51)	Course credit	63	25815
Germany	ge_po	University of Potsdam	German	43	24.89 (16-58)	14.97 (3.78)	12.5 Euros / hour	70	49604
Iceland	ic	University of Iceland	Icelandic	45	23.58 (18-30)	15.20 (2.09)	Course credit	76	56821
India	hi_iitk	Indian Institute of Technology Kanpur	Hindi	45	21.11 (19-33)	16.52 (2.89)	Course credit	78	58642
India	hi_iiith	International Institute of Information Technology Hyderabad	Hindi	53	21.41 (18-29)	17.20 (2.54)	200 Rupees / hour	88	77522
Russia	ru_mo	Higher School of Economics	Russian	49	20.67 (17-30)	13.76 (2.13)	500 Rubles / session	71	57817
Serbia	se	Universities of Belgrade and Novi Sad	Serbian	43	19.58 (18-32)	12.30 (1.67)	Course credit	65	45845
Spain	ba	Basque Center on Cognition, Brain and Language	Basque	36	NA	NA	10 Euros/hour	72	43199
Switzerland	ge_zu	University of Zurich	German	47	23.98 (18-29)	15.77 (2.83)	25 CHF / session	75	58517
UK	en_uk	University of Southampton	English	50	19.84 (18-32)	14.08 (2.94)	Course credit	80	66041
*Norway	no	University of Oslo	Norwegian	22	24.50 (19-31)	15.95 (2.03)	300 NOK gift card / session	53	19025
Taiwan	ch_t	National Taiwan Normal University	Mandarin (traditional script)	43	24.76 (20-30)	16.22 (2.00)	400 NTD / session	63	44900
*Turkey	tr	Middle East Technical University	Turkish	14	23.57 (20-27)	16.57 (1.40)	50 Turkish Lira / session	58	13865

Note: The "L2 data" for the UK sample represented L1 reading of the same 12 English texts that all other participant samples read. Samples marked with * are appended samples from the same institutions collected during Wave 1. Note that some sites paid participants for the full experimental session, including also the L1 component of the study (i.e., per session), while other sites paid participants on an hourly basis (i.e., per hour).

Materials. The English passage reading eye-tracking task consisted of 12 texts in English, compiled from the training materials for the ACCUPLACER Reading test and the English as Second Language Reading Skills test, i.e., the placement tests often taken by students in North American colleges. Each text, written in expository prose and dedicated to a historical person or natural phenomenon, came with two 4-alternative-forced-choice factual and inferential comprehension questions. Text lengths varied from 98 to 185 words (4-11 sentences). Texts and questions were presented to participants in a fixed order. Kuperman et al. (2023) report characteristics of the texts, including their length and readability. The Flesch-Kincaid grade level of readability showed that the texts were in the range expected of high-school and college-level reading (M = 10.56, SD = 2.68) and close to the range observed among advanced L2 learners of English in Crossley et al. (2011). The Coh-Metrix L2 readability score (M = 16.17, SD = 5.56) for MECO L2 texts approximated the mean values that Crossley at al. (2011) associated with readings for intermediate learners. These readability estimates thus suggest that the texts used are appropriate for our intermediate-to-advanced sample of English L2 readers. For further details, we refer readers to Kuperman et al. (2023). .

Additional questionnaires and tests. Participants in all samples completed the same series of tests and questionnaires. This series included a battery assessing component skills in English (see below), and a non-verbal intelligence test (the Culture Fair Test-3 CFT20, Subset 3 Matrices, short version, Form A, timed at 3 minutes, Weiß, 2006). Further, an abridged version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) collected basic demographic and linguistic information about the participants' use of and proficiency in L1 and English: E.g., the participants' age and years of education, age when learning English began, and self-ratings of their proficiency in their L1 and in English. Please note that, for simplicity, we designate English as L2 for all samples (except the UK) even though English may be L3 or LX for some samples or some participants. The full information

collected through the questionnaire, including the age of acquisition and proficiency in each language, is available through the project's repository (see below).

English reading comprehension and fluency are demonstrably contingent on the reader's mastery of component skills of English language and reading proficiency (see reviews by Gillon, 2017; Jeon & Yamashita, 2014; Koda, 2005; Schmitt, 2008; and Vandergrift, 2007, among others). The MECO L2 project taps into some of those component skills through administering an additional battery of six tests of individual differences. Test (1) was the Spelling Recognition test (adapted from Andrews & Hersch, 2010): In this test items are presented in a list and participants need to decide for each whether it is a correctly spelled word in English or not (i.e., mark each item as 'correct' or 'incorrect'). Half of the items are correctly spelled and the other half include spelling errors (e.g. seperate, benafit). Test (2) was a Vocabulary Knowledge test based on word recognition with multiple-choice questions (adapted from Nation & Beglar, 2007). For this test of the receptive knowledge of English, words are selected from a frequency-ranked list of 14,000 English lemmas and ten items are chosen from each 1000 words in the ranked list to represent the respective frequency band. The test consists of a series of questions where a target word is embedded in a short non-defining context and participants need to choose its correct definition from four options. Test (3) consisted of the assessment of motivation to excel in the task (using the Student Opinion Scale, SOS, questionnaire; Thelk et al., 2009). The SOS includes 10 statements that participants are asked to rank from '1=Strongly Disagree' to '5=Strongly Agree' according to how they feel about each of them in relation to completing the current study. Test (4) is the lexical knowledge test LexTALE with yes/no decisions (Lemhöfer & Boersma, 2012). It is an untimed lexical decision task, consisting of 60 trials: 40 words and 20 pseudowords. Tests (5-6) come from the TOWRE test of reading efficiency (TOWRE-II; Torgesen et al., 2012), with one subtest for word naming (Sight Word Efficiency) and one subtest for pseudoword naming (Phonemic Decoding Efficiency). In each subtest participants are required to read aloud as many items as possible from a list of words/pseudowords within a 45-

second time limit. Altogether, these tests tap into the reader's ability to associate sounds and letters of the written word (decoding); ability for word identification; spelling ability as a measure of orthographic learning and knowledge; vocabulary knowledge as a central ability for word recognition and comprehension; and an extra-linguistic motivational component. This battery of tests was again identical to that collected in the first wave of the project: We thus do not repeat the full details regarding the scoring and administration of tests as these are fully available in Kuperman et al. (2023), Supplementary Materials S2. The tasks in the battery were administered in the fixed order (1)-(6), after the completion of the main eye-tracking passage reading task. Tasks (1)-(3) were administered using an in-house web-based platform; task (4) was administered through the LexTALE website (http://www.lextale.com/); and tasks (5) and (6) were administered in the standard pencil-and-paper version.

Due to different reasons – including administration errors, connectivity issues, and copyright restrictions due to which a few sites opted out of the CFT-20 test or TOWRE – some participants in the MECO L2 sample do not have complete data in all verbal and background skill tests. We report details regarding the number of missing values in each test in Supplementary Materials S1.

Procedure Participants were tested individually. In the beginning of the experimental session, participants signed a consent form and completed the LEAP-Q questionnaire. Then, participants completed an L1 reading task where they read 12 texts in their L1 silently for comprehension while their eye-movements were recorded, followed by four yes/no comprehension questions after each text. Then, participants proceeded to a skill-test battery in L1, which included the CFT-20 and other tests of individual differences in L1. With the exception of the CFT-20 task, data collected during these stages of the experiment is reported elsewhere. The current paper reports data when participants proceeded to the English component of the project, i.e., the task of silently reading 12 texts in English for comprehension, while their eye-movements were recorded. The reading task was followed by the battery of English individual differences tasks described above. The duration of the L2 eye-tracking

reading task was roughly 20-30 minutes, and the individual differences battery took up to 30 minutes[2]. The entire session lasted no more than two hours, and breaks were provided as requested[3]. All data were collected by research assistants trained in eye-tracking data collection according to the protocols of their labs.

Apparatus As outlined in Kuperman et al.'s (2023) Method section, to record eye movements during reading, all participating laboratories used an EyeLink Eye-tracker (SR Research, Kanata, Ontario, Canada). Labs had one of the Portable Duo, EyeLink II, 1000 or 1000+ models. A sampling rate of 1000 Hz was used in all sites but Serbia, where the EyeLink II was used with a sampling rate of 500 Hz. All sites used the same experimental procedure programmed in the Experiment Builder software (SR Research). A chin rest was used to minimize head movements. Calibration was performed using a series of nine fixed targets distributed around the display, followed by a 9-point accuracy test to validate eye position. Stimuli were viewed binocularly but eye-movement data were analyzed from only the self-reported dominant eye (the right eye in most participants). Before presenting the trial stimuli in the English eye-tracking reading task, a dot appeared on the monitor screen, slightly to the left of the first word in the passage. Once the participant had fixated on it, the trial began. This drift check took place at the beginning of each trial, and calibration was monitored by the experimenter throughout the task and was redone if necessary. Each of the 12 texts appeared on a separate screen. Participants were instructed to read the passages silently for comprehension and press the space bar when their reading of a passage was completed. A mono-spaced font (Consolas) was used with a size generally ranging from 20 to 22 points (given variation in screen size and resolution at different testing sites) and 1.5 line spacing. In

[2]The UK sample completed these tests as part of their L1 individual-differences battery, and so their testing session was shorter than in the other sites.

[3] There was one exception to the described testing order. For logistic reasons, in Serbia, participants completed two separate testing sessions: The first consisted of the L2 (English) battery, including the eye-tracking L2 data collection, skills of individual differences, CFT, and LEAP-Q; and the second consisted of the L1 eye-tracking reading task and individual differences in L1.

accordance with their local experimental setup, the German site in Zurich used a smaller font size of 10 with a lower resolution of 1280 x 1024. The refresh rate was set to 60 Hz at all sites. For further specifications of the screen, font size, presentation settings, and apparatus at each participating site, see Supplementary Materials S2. Each text was followed by two multiple-choice comprehension questions, shown on a separate screen one after another. Participants responded by choosing their answers using the number keys (1-4).

Data editing and cleaning. Data editing and cleaning. The *popEye* software was used to pre-process the eye-tracking data (implemented in R, version 0.8.1, Schroeder, 2019). During this process, fixations are automatically corrected on the vertical axis and assigned to lines. In the current Wave 2 of MECO, the "slice" algorithm was used, because it was shown to provide a substantial boost in assignment accuracy compared to the baseline algorithm used for Wave 1 (Glandorf & Schroeder, 2021). However, in the two appended samples that added participants to a Wave 1 sample (i.e., in Turkey and Norway) the baseline algorithm was used to maintain consistency within a site. Following this automatic procedure, members of the research team visually inspected the output of the software and assessed the quality of the resulting data. The assessment consisted of detecting texts in which fixations and text lines were misaligned (due to poor calibration). Such texts were removed from the data pool, as were participants with less than 5 (out of 12) usable texts with the high-quality eye movement record. Table 1 reports the percent of remaining texts and word tokens (interest areas) after this data cleaning.

For the purposes of reliability and descriptive analyses below, further data cleaning involved removing data points that showed very short (< 80 ms) first fixations, which are unlikely to provide sufficient time to complete visual uptake (see Warren et al., 2009), or very long total fixation times (top 1% of the participant-specific distribution, all exceeding 3s on the word). A total of 15,400 data points (2.0% of total) were removed, between 1.3 and 2.3% per language. Off-screen looks were incorporated

in the passage-level variables (e.g., reading rate) but not in the word-level eye-tracking variables (see details on variables used, below).

Data availability. As with the previous release of MECO reported in Kuperman et al. (2023), the current Wave 2 release of MECO L2 includes full interest-area reports from usable participants and trials, as well as passage- and sentence-level summaries. Also included are full data from individual differences tests in L2, the non-verbal IQ test, and the background questionnaire. Please refer to the project's repository page at https://osf.io/g5mxb/?view_only=bf4698c176d446cd9d7a8215856f26f0 for the full materials, the analysis code, and data. Note that the data provided in the MECO L2 Wave 2 repository can be easily aggregated with the Wave 1 data (i.e., data structures are similar), previously reported and made available in Kuperman et al. (2023).

[**Note to editor and reviewers**: This is a view-only link to a repository that includes MECO L2 Wave 2 data only, for reviewing purposes. Upon publication, the Wave 2 data will be moved to the general MECO L2 repository. Also note that the full Wave 2 release of the L1 data from the same samples of participants will be soon available through the MECO L1 repository, which will enable L1-L2 within-participant comparisons].

Reading variables. A number of eye-movement variables are considered as measures of reading fluency (both in L1 and L2). In our description below, we follow closely on Kuperman et al.'s (2023) variable definitions. The word-level variables include *skipping* (a binary index of whether the word was fixated upon at least once during the entire text reading, labeled as *skip*[4]). For words that were fixated at least once, the following variables were defined: *first fixation duration* (the duration of the first fixation

[4] The data we make available also include a variable (firstrun.skip) for whether the word was skipped during the first reading pass. While this variable finds more use in word and sentence reading, it is more problematic in studies of text reading. Quite often, readers begin with inspecting the length of the text to be read and so the first few fixations may land towards the middle or the end of a text passage: under a traditional definition, most words in such scenario would be considered skipped, leading to massive data loss for the fixation analysis.

landing on the word, *firstfix.dur*); *gaze duration* (the summed duration of fixations on the word in the first pass, i.e., before the gaze leaves it for the first time, *firstrun.dur*); *total fixation duration* (the summed duration of all fixations on the word, *dur*); *number of fixations* on the word (*nfix*); *refixation* (a binary index of whether a word elicited more than one fixation in the first pass, *refix*); *regression-in* (a binary index of whether the gaze returned to the word after inspecting further textual material, i.e., to the right of the word in left-to-right orthographies, *reg.in*); and *re-reading* (a binary index of whether the word elicited fixations after the first pass, i.e., after the gaze left the word for the first time, *reread*[5]). See Inhoff and Radach (1998), Rayner (1998), and Godfroid (2020) for a detailed discussion of these variables. At the participant level, the following measures of fluency were defined: reading rate (in words per minute, *rate*), and mean word-level variables (e.g., participant's mean skipping rate, mean first fixation duration, etc.). Sentence and passage reading times, as well as the number of fixations, skips, and regressions per sentence and passage can be found in the sentence and passage-level reports, respectively, in the project's data repository. Finally, we gauged *comprehension accuracy* as the percent of correct responses to all 24 questions (*acc*). Computed variables are identical and backward compatible with variables in the first wave of the project, enabling future analyses on the aggregate data across Wave 1 and Wave 2 sites.

Tests of individual differences provide the following set of dependent variables: Scores from the CFT test of non-verbal intelligence (*cft)*, as well as scores on tests of spelling (*spelling)*, vocabulary knowledge (*vocabulary*[6]*),* motivation (*motivation)*, LexTALE (*lextale)*, Sight Word Efficiency (*towre*: *swe)*

[5] An alternative measure of rereading can be computed using the MECO data to examine not just whether a word was reread, but how long rereading took. This can be done by subtracting gaze duration from total reading time.
[6] As discussed at length in Kuperman et al., 2023, two measures were computed based on the vocabulary knowledge test: One based on data across all available blocks ("thousands" 2-10), and another based on responses in earlier blocks only ("thousands" 2-5). As this is an adaptive task, with stopping rules at the end of each block ("thousand"), many participants had little to no data in later blocks. The adapted measure from thousands 2-5 focuses on parts of the test where most participants have substantial data, and indeed was shown to be more reliable in both Kuperman et al., 2023, as well as our data (see *reliability estimates*, below). Similar to Kuperman et

and Phonemic Decoding Efficiency (*towre: pde*; see details regarding the scoring of individual differences tests in Kuperman et al., 2023, Supplementary Materials S2).

Results

Below, we first report reliability analyses of eye-tracking data and individual differences tests in the current Wave 2 and compare those estimates against reliability previously observed in Wave 1 of the MECO L2 project. We follow with presenting the descriptive statistics of the Wave 2 data and correlational analyses that pit eye movement measures against themselves and against the skill test scores. In all sets of analyses, we follow the analytical procedure of Kuperman et al. (2023), for comparability.

Reliability estimates

Eye-tracking Data

The split-half reliability at the participant-level for an eye-tracking measure reveals how stable that measure is given individual differences between participants. This reliability metric is the correlation between mean values for 'odd' and 'even' words within a participant. Specifically, we would calculate mean values for, say, gaze duration for words (i.e., interest areas) 1, 3, 5, etc. and words 2, 4, 6, etc. for each participant. Reliability can then be estimated as the correlation between the mean values for 'odd' and 'even' words across all participants in the sample. The participant-level reliability for reading rate was estimated using an Intra-class Correlation Coefficient (ICC), measuring the degree of agreement in reading rate estimates across the 12 texts. In addition, a word-level reliability estimation was done at the *word token*-level. This reliability is of interest for studies of the effect that word properties have on eye movements. For each word token in the MECO texts, mean values were calculated for each eye

al., 2023, we thus use the adapted measure throughout this paper. Both measures are available in the project's repository for interested users.

movement measure for "odd" and "even" *participants* separately. The resulting two sets of values were correlated across all word tokens to form a reliability estimate.

Supplementary Materials S3 and S4 provide a full report of the two types of reliability estimates (i.e., participant-level and word token-level). Similar to reliability analyses in Wave 1, all eye movement measures in Wave 2 demonstrate an extremely high reliability of eye-tracking measures at the participant level (all Spearman-Brown corrected reliability estimates > 0.90). In line with Staub (2021), this finding indicates that the eye-movement measures faithfully reflect individual differences in English proficiency. As expected and in line with Kuperman et al., 2023, reliability at the word token-level was considerably lower. Still, the average Spearman-Brown corrected reliability estimates, aggregated across sites and measures, were in the moderate range (mean r = 0.68, median 0.71), as were the reliability estimates for most measures and samples. Again, reliability levels found in the Wave 2 data were highly comparable to those in MECO's Wave 1 (e.g., Spearman–Brown corrected reliability estimates r > 0.94 at the participant level and r > 0.6 at the item level in Kuperman et al., 2023) as well as the GECO database (between 0.6 and 0.9 in Cop et al., 2017).

Tests of Component Skills and Comprehension Accuracy

In addition to eye-movement measures, we calculated the reliability for scores in the online battery of English skill tests (spelling, vocabulary, and motivation[7]), as well as for comprehension accuracy in the passage reading task. For comprehension, spelling, and motivation we calculated both split-half reliability as well as Cronbach's alpha. For the vocabulary knowledge task, we only calculated split-half because of the adaptive nature of this task, which means that different participants have data from

[7] Reliability could not be calculated for TOWRE as the test is based on a single word and a single pseudoword list. TOWRE scores are expected to be highly reliable, as reflected in previous reports of high test-retest reliability estimates (Torgesen et al., 2012). Previous reports also establish LexTALE as a reliable measure in L2-English participants (Lemhöfer & Broersma, 2012).

different trials (see design details in Kuperman et al., 2023). Reliability estimates were calculated on the aggregated dataset (not broken down by site), as procedural differences across sites are not expected to have an impact on the data quality in these tests. The estimates are provided in Supplementary Materials S5. Unsurprisingly, these estimates were highly similar to the ones reported in Kuperman et al. (2023) – this is expected given the highly similar nature of participants in the two waves of the project. Specifically, reliability estimates for the four tests - spelling, motivation, vocabulary, and comprehension - were reasonable, with split-half estimates ranging from 0.64 to 0.75 and Cronbach's alpha values of 0.61 to 0.73. In sum, MECO L2 data on reading fluency and comprehension, as well as the test scores in component skills of English reading, show acceptable to high levels of reliability, making the data eligible for a meaningful inferential analysis.

Descriptive and correlation analyses

Figure 1 visualizes means and standard deviations of eye movement measures and comprehension accuracy by language sample. These estimates were obtained by first calculating the means for these variables by participants and then aggregating those by-participant means. Detailed data summaries, organized by variable and sample, are provided in the project's repository. Figure 2 further shows the means and standard deviations of scores in the available measures of individual differences (including tests of component skills and non-verbal intelligence).

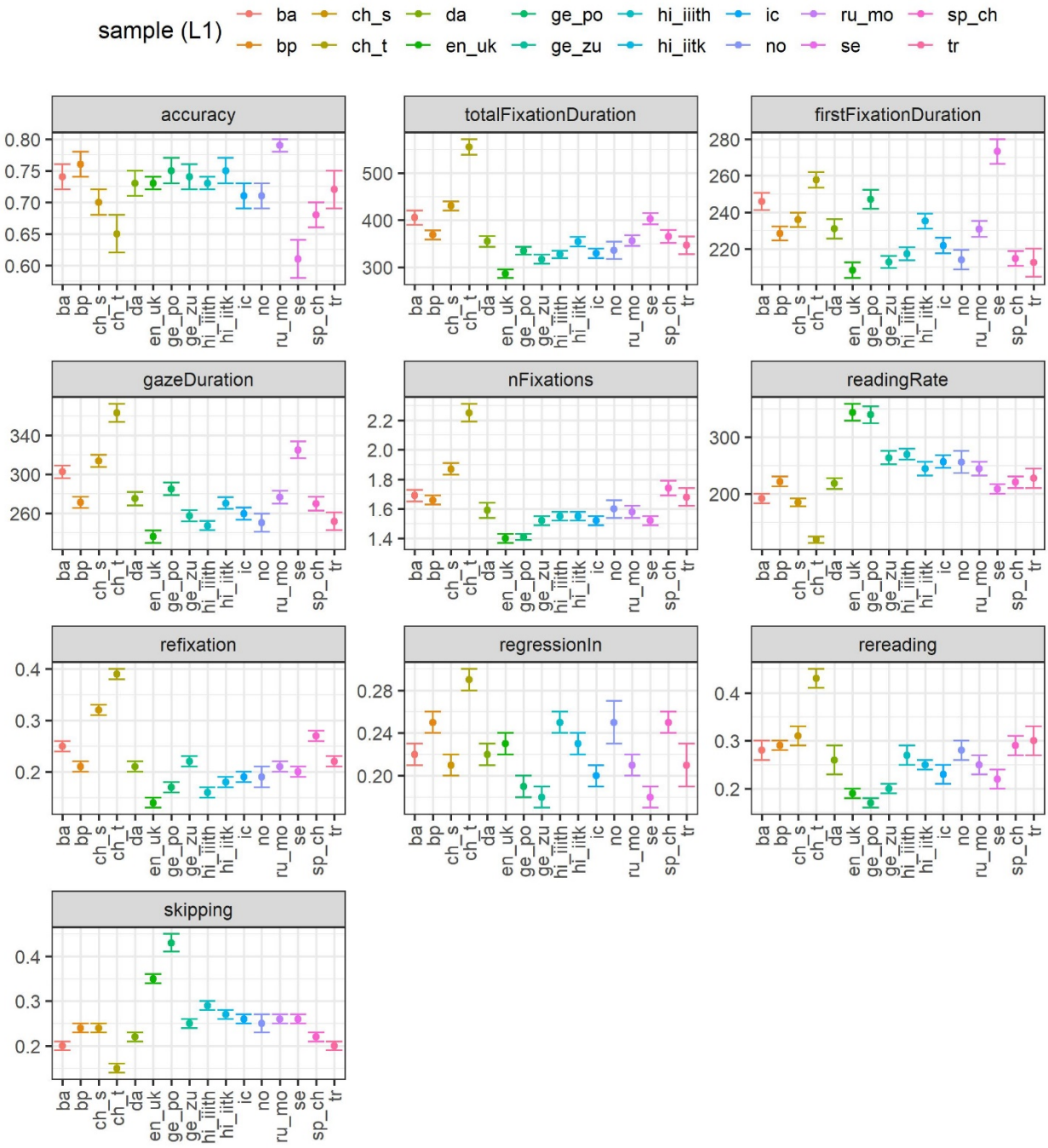

Figure 1. Means of measures from the eye-tracking task across samples. Error bars stand for ± 1 SE. accuracy: percent answers correct; *accuracy*: comprehension accuracy; *nFixation*: number of fixations; *refixation*: likelihood of second fixation on the word; *regressionIn*: regression rate; *rereading*: likelihood of second pass; *skipping*: skipping rate. *ba:* Basque*; bp* – Brazilian Portuguese; *ch_s* – Chinese simplified*; ch_t* – Chinese traditional*; da* – Danish*; en_uk* – English (UK sample)*; ge_po* - German (Potsdam sample)*; ge_zu* - German (Zurich sample)*; hi_iiith* - Hindi (Hyderabad sample)*; hi_iitk* - Hindi (Kanpur sample)*; ic* – Icelandic*; no* – Norwegian*; ru_mo* – Russian (Moscow sample)*; se* – Serbian*; sp_ch* – Spanish (Chile sample)*; tr* - Turkish.

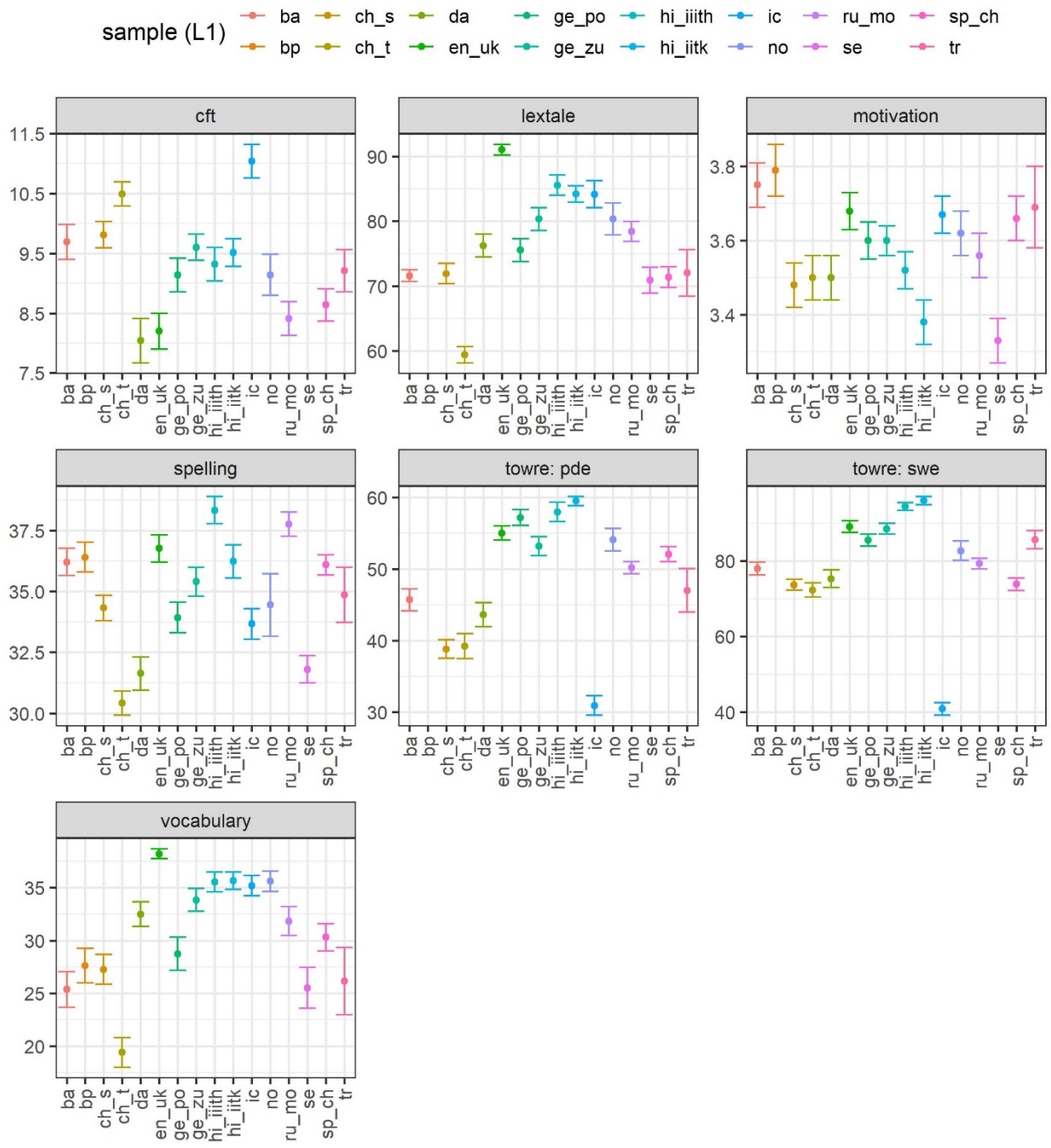

Figure 2. Means of measures of individual differences of English proficiency across samples. Error bars stand for ± 1 SE. cft: score in the CFT test; towre: pde: TOWRE, Phonemic Decoding Efficiency subtest (pseudoword naming); towre: swe: TOWRE, Sight Word Efficiency subtest (word naming); vocabulary: vocabulary knowledge (Groups 2-5). *ba:* Basque*; ch_s* – Chinese simplified*; ch_t* – Chinese traditional*; bp* – Brazilian Portuguese; *da* – Danish*; en_uk* – English (UK sample)*; ge_po* - German (Potsdam sample)*; ge_zu* - German (Zurich sample)*; hi_iiith* - Hindi (Hyderabad sample)*; hi_iitk* - Hindi (Kanpur sample)*; ic* – Icelandic*; no* – Norwegian*; ru_mo* – Russian (Moscow sample)*; se* – Serbian*; sp_ch* – Spanish (Chile sample)*; tr* - Turkish.

A comprehensive analysis of these descriptive patterns and the cross-site differences and similarities that emerge from Figures 1 and 2 is beyond the scope of this paper. However, we do want to highlight a few important observations that again establish the quality of the MECO Wave 2 data. First, we note that there was substantial similarity across samples in terms of comprehension accuracy: Specifically, 11 out of the 16 samples showed comprehension accuracy in a similar range of 70% to 75% (a range far from ceiling performance). This picture is very much in line with MECO Wave 1 data (where 8 out of the previous 12 samples showed comprehension accuracy in a similar range). In contrast, and again in line with Kuperman et al. (2023), there was much more variability in oculomotor measures of fluency. This is true both within the different L2 samples, but also, most notably, in how estimates of oculomotor measures in the English L1 sample (i.e., in the UK) stand out among the sites where participants are L2 readers of English. Visual inspection of Figure 1 demonstrates that the English L1 readers (en_uk sample) had a faster reading rate, shorter and fewer fixations, a higher skipping rate, and a lower likelihood of refixations or re-reading compared to most L2 samples. There were some L2 samples (e.g., two samples of German speakers) that showed a similar or even more extreme mean values than English L1 in some eye movement measures. Yet no L2 sample showed as consistent a contrast with the majority of L2 samples as the sample of L1 English speakers. See also Siegelman et al. (2023) for evidence of the comprehension-fluency contrast in terms of L1-L2 differences and similarities.

Figure 2 further presents extensive variability in performance on tests of component skills across the sites, with the English L1 (en_UK) sample showing generally higher performance than other sites in these tests, with further expected variability among L2 sites. We further note that in both Figure 1 (i.e., eye-movement measures) and Figure 2 (i.e., component skill tests), it is hard to find, from a cursory look, a clear linguistic factor that maps directly into the observed behavioral similarities and differences. Taken together, these observations replicate those from the Wave 1 data and open exciting avenues for systematic analyses of the determinants of oculomotor measures of L2 reading given various properties

of participants' L1 across the various language backgrounds and their component skills (see more in the General Discussion).

Lastly, we computed the correlations between eye-movement measures, accuracy, reading rate, and all individual differences tests on the aggregated data set of participants from all Wave 2 samples (N = 661; Table 2). Correlational analyses like this speaks to some of the central questions in second language acquisition research, e.g., does reading fluency correlate with reading comprehension; and does individual variability in component skills of reading influence reading comprehension and fluency. They also help answer methodological questions about the inter-sample differences, potentially driven by variability in the non-verbal IQ and motivation to perform well in the task.

We again replicated four main correlational findings in Kuperman et al. (2023): (1) there were substantial correlations between the various eye-movement measures; (2) there were only weak correlations between comprehension accuracy and the oculomotor reading measures ($|r|$ between 0.03 and 0.35); (3) individuals with higher performance in the English component skill tests had more efficient eye-movement reading patterns (i.e., more skips, fewer and shorter fixations); and (4) CFT and motivation were only weakly correlated with eye-movement measures ($|r| \leq 0.16$). These expected correlational patterns suggest, in line with Kuperman et al.'s (2023) data, that reading fluency (gauged by eye movements) and comprehension are only weakly related, proficiency in component skills of reading influences reading behavior, and the inter-sample variability in IQ and motivation did not strongly affect eye movement patterns. The alignment with Kuperman et al.'s report from MECO Wave 1 validate the current extension of the MECO L2 data.

Table 2. Correlation table for reading measures (data aggregated across samples, N = 661). Values above the diagonal show Pearson correlation coefficients; values below the diagonal show *p* values (p-value shown as 0 stands for $p < .001$), and significant correlations ($p < .05$) appear in bold text.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1) skipping rate		**-0.68**	**-0.59**	**-0.30**	**-0.16**	**-0.43**	**-0.7**	**-0.56**	**0.86**	**0.16**	**0.27**	**0.32**	0.07	**0.33**	**0.25**	**0.37**	**-0.09**
2) n fixations	0		**0.84**	**0.62**	**0.17**	**0.47**	**0.8**	**0.91**	**-0.78**	**-0.15**	**-0.22**	**-0.40**	**-0.09**	**-0.31**	**-0.23**	**-0.39**	**0.14**
3) total fixation time	0	0		**0.42**	**0.66**	**0.84**	**0.72**	**0.71**	**-0.84**	**-0.27**	**-0.31**	**-0.49**	**-0.14**	**-0.45**	**-0.24**	**-0.40**	**0.10**
4) regression rate	0	0	0		-0.08	0.02	**0.25**	**0.75**	**-0.39**	0.05	0.05	**-0.09**	0.03	0.03	-0.05	**-0.11**	0.07
5) first fix duration	0	0	0	0.054		**0.87**	**0.23**	0.07	**-0.50**	**-0.27**	**-0.28**	**-0.35**	**-0.12**	**-0.35**	**-0.13**	**-0.19**	-0.03
6) gaze duration	0	0	0	0.649	0		**0.64**	**0.23**	**-0.70**	**-0.35**	**-0.39**	**-0.48**	**-0.16**	**-0.49**	**-0.24**	**-0.37**	0.03
7) refixation	0	0	0	0	0	0		**0.50**	**-0.74**	**-0.26**	**-0.38**	**-0.45**	**-0.13**	**-0.45**	**-0.29**	**-0.44**	**0.09**
8) rereading	0	0	0	0	0.070	0	0		**-0.67**	-0.03	-0.05	**-0.26**	-0.04	**-0.14**	**-0.15**	**-0.27**	**0.13**
9) reading rate	0	0	0	0	0	0	0	0		**0.22**	**0.31**	**0.43**	**0.11**	**0.43**	**0.26**	**0.40**	**-0.09**
10) accuracy	0	0	0	0.181	0	0	0	0.512	0		**0.39**	**0.40**	**0.24**	**0.37**	**0.20**	**0.25**	**0.13**
11) spelling	0	0	0	0.181	0	0	0	0.208	0	0		**0.37**	**0.20**	**0.45**	**0.27**	**0.42**	0.04
12) vocabulary	0	0	0	0.027	0	0	0	0	0	0	0		**0.21**	**0.57**	**0.15**	**0.29**	0.00
13) motivation	0.085	0.020	0.001	0.505	0.002	0	0.001	0.310	0.006	0	0	0		**0.16**	0.01	0.03	0.06
14) LexTALE	0	0	0	0.533	0	0	0	0.001	0	0	0	0	0		**0.22**	**0.26**	**0.10**
15) towre: swe	0	0	0	0.25	0.002	0	0	0	0	0	0	0.001	0.872	0		**0.70**	**-0.14**
16) towre: pde	0	0	0	0.01	0	0	0	0	0	0	0	0	0.497	0	0		**-0.12**
17) cft	0.038	0.001	0.024	0.096	0.510	0.522	0.025	0.002	0.032	0.002	0.403	0.992	0.164	0.024	0.001	0.006	

Notes: *n fixations*: number of fixations; *first fix duration*: first fixation duration; *refixation*: likelihood of second fixation on the word; *rereading*: likelihood of second pass; *accuracy*: comprehension accuracy; *vocabulary*: vocabulary knowledge (Groups 2–5); *towre: swe*: TOWRE, Sight Word Efficiency subtest (word naming); *towre: pde*: TOWRE, Phonemic Decoding Efficiency subtest (pseudoword naming); *cft*: score in the CFT test.

General Discussion

This paper reports an expansion of the English as L2 reading component of the Multilingual Eye Movements Corpus (MECO L2; Kuperman et al., 2023). This Wave 2 adds eye-tracking data on text reading in English, as well as scores from component skills of English proficiency, from 16 laboratories worldwide and 13 unique L1s. Tests of component skills of the English proficiency include spelling, vocabulary knowledge, lexical decision, sight word efficiency, and phonemic decoding efficiency, as well as non-verbal intelligence and motivation to excel in the task. Questionnaires offer additional insight into demographics of participants as well as their background and use of their L1 and English. Some of the samples in the present report increase the size of the samples collected in Wave 1 ("appended samples" in Turkish and Norwegian), some represent the languages already found in Wave 1 but recruit participants from a different country or university ("replication samples" in German, Russian, Spanish, English), and the majority of the samples come from L1s new to the MECO project.

Jointly with Wave 1, the English-reading component of the MECO corpus currently encompasses readers of English with 19 distinct L1 language backgrounds, includes English as L1 (Canada, UK) and, primarily, L2 readers of English. The language backgrounds of the readers of English in the MECO project incorporate a large typological and genetic variety of languages (Basque, Indo-European, Semitic, Sino-Tibetan, and Turkic language families) and writing systems (e.g., Chinese simplified logographic, Hebrew abjad, Hindi abugida, and several alphabets). Participants in the current MECO L2 component also contributed data in their L1, enabling within-participant comparisons: The L1 data are reported elsewhere. We note that MECO is an evolving and ongoing project, and its future releases (e.g., Wave 3) plan to further enrich this data resource with behavioral samples of L1 and L2 reading from readers of diverse languages and writing systems.

As is the case with any data resource, MECO has its limitations. Its samples are of relatively small size (around 50 participants), which limits cross-linguistic comparisons at the participant level. The battery of component skills of reading is lacking tests of several skills that are known to strongly contribute to L2 reading proficiency (e.g., L2 listening comprehension). Also, since availability of tests of individual differences varies drastically across languages, we do not administer a battery of tests for proficiency in L1, which is a factor of major influence on L2 reading proficiency.

Analyses in this paper demonstrate very high reliability of eye-movement data at the participant level, and moderate-to-good reliability of eye-movement data at the word-token level, comprehension accuracy, and all tests of component skills. Thus, the data have adequate quality both for group-level comparisons and for the study of individual differences. Another validation of the quality of the Wave 2 data comes from the observed correlation patterns, which match those uncovered in Wave 1 of the MECO project (Kuperman et al., 2023). Among other findings, we found that the L1-L2 differences and the overall variability in English reading comprehension are minor relative to L1-L2 differences and variability in all measures related to reading fluency. While L1 English speakers demonstrate comprehension accuracy comparable to that in most L2 samples, they were much more fluent (shorter reading times, faster reading rate, etc.) than the L2 counterparts. This dissociation between reading comprehension and fluency, observed in Kuperman et al. (2023), is a fruitful topic for future research.

More broadly, the goal of the current paper is simply to establish the reliability and quality of the MECO Wave 2 data so that follow-up analyses can mine it in future studies into different facets of L2 reading. Multiple interesting avenues include (i) a comparison of English reading performance between speakers of the same language versus speakers of different languages (e.g., does a specific L1 background has a footprint that makes, say, German readers of English more similar to one another than to English readers with other non-native backgrounds?); (ii) the effect of the degree of similarity between the L1 background of the reader and English on a reader's reading comprehension and fluency,

and (iii) the determinants of (various facets of) English proficiency, and the relative contribution of skill tests (e.g., spelling, vocabulary knowledge), one's L1 background and its similarity to English, and other participant-level characteristics. With the full MECO L2 data made freely available in the spirit of Open Science, we hope that these and many other questions are investigated by the community of researchers of L2 reading.

Acknowledgements

Research reported in this publication was supported by the following grants: The Social Sciences and Humanities Research Council of Canada Partnered Research Training Grant, 895-2016-1008 (PI: G. Libben); the Social Sciences and Humanities Research Council of Canada (SSHRC) Insight Grant, 435-2021-0657; the Canada Research Chair (Tier 2; PI: V. Kuperman); the German Federal Ministry of Education and Research (BMBF), 01| S20043 (PI: L. A. Jäger); the National Council for Scientific and Technological Development of Brazil (Conselho Nacional de Desenvolvimento Científico e Tecnológico) Project 316036/2021-8 (PI: R. Rothe-Neves); the Obel Family Foundation, Research Equipment Grant to Aalborg University, 2017 (PI: H. B. S. Knudsen); the UKRI Economic and Social Research Council South Coast Doctoral Training Partnership, ES/P000673/1; Project Fondecyt Regular by the National Research and Development Agency (ANID-CHILE), 1201440 (PI: R. Ibáñez Orellana). Project Fondecyt Postdoctorado by the National Research and Development Agency (ANID-CHILE), 3210252 (PI: A. Santana Covarrubias); the "Chinese Language and Technology Center, National Taiwan Normal University" within the Higher Education Sprout Project framework by the Ministry of Education in Taiwan (PI: Y. T. Sung); the Basic Research Program at the National Research University Higher School of Economics (HSE University); the Ministry of Science, Technological Development and Innovation of the Republic of Serbia; the Israel Science Foundation (ISF) Grant, project 1034/23 (PI: N. Siegelman), and by an Azrieli Early Career Faculty Fellowship (PI: N. Siegelman).

We wish to thank the following individuals: Yaqian Bao, Itziar Basterra, Isidora Damjanović, Ainhoa Eguiguren, Amets Esnal, Brianna Griska-Macphee, Nora Hollenstein, Chia En Hsieh, Alexandra Jackson, Nadia Lana, Jolie Luk, Sriya Ravula, Evonne Syed, and Lucy Thomas.

References

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139*(2), 299.

Berzak, Y., Nakamura, C., Smith, A., Weng, E., Katz, B., Flynn, S., & Levy, R. (2022). CELER: A 365-participant corpus of eye movements in L1 and L2 English reading. *Open Mind, 6*, 41-50.

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of E*ducational *P*sychology, 97(4), 580.

Brysbaert, M., & Drieghe, D. (2024). The use of eye movement corpora in vocabulary research. *Research Methods in Applied Linguistics, 3*(1), 100093.

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods, 49*, 602-615.

Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: comparison of readability formulas. *Reading in a Foreign Language, 23*, 84–101.

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences, 9*(3), 33.

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. Reading and *W*riting, 12, 1-30.

Gillon, G. T. (2017). *Phonological awareness: From research to practice*. Guilford Publications.

Glandorf, D., & Schroeder, S. (2021). Slice: an algorithm to assign fixations in multi-line texts. *Procedia Computer Science, 192,* 2971-2979.

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York: Routledge.

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. Bilingualism: *Language and Cognition, 23*(2), 283-294.

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. *Eye guidance in reading and scene perception*, 29-53.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160-212.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach.* New York: Cambridge University Press.

Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., ... & Usal, K. A. (2023). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition, 45*(1), 3-37.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325-343.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605-621.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967.

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin, 140*(2), 409.

Nation, P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372-422.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363.

Schroeder, S. (2019). popEye-An integrated R package to analyse eye movement data from reading experiments. *Journal of Eye Movement Research, 12*(7).

Siegelman, N., Elgort, I., Brysbaert, M., Agrawal, N., Amenta, S., Arsenijević Mijalković, J., ... & Kuperman, V. (2023). Rethinking First Language–Second Language Similarities and Differences in English Proficiency: Insights from the ENglish Reading Online (ENRO) Project. *Language Learning.*

Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H. D., Alexeeva, S., Amenta, S., ... & Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods, 54*(6), 2843-2863.

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language, 116*, 104190.

Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2023). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods, 55*(6), 2743-2763.

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *The Journal of General Education,* 129-151.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). Test of word reading efficiency–second edition (TOWRE-2). *Austin, TX: Pro-Ed*.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40(3)*, 191.

Warren, T., White, S. J., & Reichle, E. D. (2009). Investigating the causes of wrap-up effects: Evidence from eye movements and E–Z Reader. *Cognition, 111*(1), 132-137.

Weiß, R. H. (2006). *Grundintelligenzskala 2 mit Wortschatztest and Zahlenfolgetest* [Basic intelligence scale 2 with vocabulary knowledge test and sequential nu