Manuscript Running Header: The ENglish Reading Online (ENRO) Project

Article Type *(delete all but one as appropriate)***:** EMPIRICAL STUDY

Manuscript Title: Re-thinking L1-L2 Similarities and Differences in English Proficiency: Insights from the ENglish Reading Online (ENRO) Project

Author(s) Noam Siegelman[1,2], Irina Elgort[3], Marc Brysbaert[4], Niket Agrawal[5], Simona Amenta[6], Jasmina Arsenijević Mijalković[7], Christine S. Chang[8], Daria Chernova[9], Fabienne Chetail[10], A.J. Benjamin Clarke[11], Alain Content[10], Davide Crepaldi[12], Nastag Davaabold[13], Shurentsetseg Delgersuren[13], Avital Deutsch[1], Veronika Dibrova[14], Denis Drieghe[15], Dušica Filipović Đurđević[7,16], Brittany Finch[17], Ram Frost[1], Carolina A. Gattei[18], Esther Geva[19], Aline Godfroid[17], Lindsay Griener[20], Esteban Hernández-Rivera[21], Anastasia Ivanenko[22], Juhani Järvikivi[20], Lea Kawaletz[23], Anurag Khare[5], Jun Ren Lee[8], Charlotte E. Lee[15], Christina Manouilidou[24], Marco Marelli[6], Timur Mashanlo[14], Ksenija Mišić[7], Koji Miwa[25], Pauline Palma[21], Ingo Plag[23], Zoya Rezanova[14], Enkhzaya Riimed[13], Jay Rueckl[2,26], Sascha Schroeder[27], Irina A. Sekerina[28], Diego E. Shalom[29,18], Natalia Slioussar[22,9], Neža Marija Slosar[24], Vanessa Taler[30,31], Kim Thériault[30], Debra Titone[21], Odonchimeg Tumee[13], Ross van de Wetering[3], Ark Verma[5], Anna Fiona Weiss[32], Denise Hsien Wu[8], Victor Kuperman[33]

Author Affiliations [1]The Hebrew University of Jerusalem [2]Haskins Laboratories [3]Victoria University of Wellington [4]Ghent University [5]Indian Institute of Technology – Kanpur [6]University of Milano-Bicocca [7]University of Belgrade [8]National Taiwan Normal University [9]Saint Petersburg State University [10]Université Libre de Bruxelles [11]Thammasat University [12]International School for Advanced Studies (SISSA) [13]Khovd Branch of National University of Mongolia [14]Tomsk State University[15]University of Southampton [16]University of Novi Sad [17]Michigan State University [18]Universidad Torcuato Di Tella [19]University of Toronto [20]University of Alberta [21]McGill University [22]Higher School of Economics (HSE) Moscow [23]Heinrich-Heine-Universität Düsseldorf [24]University of Ljubljana [25]Nagoya University [26]University of Connecticut [27]University of Goettingen [28]City University of New York [29]Universidad de Buenos Aires [30]University of Ottawa [31]Bruyère Research Institute [32]Catholic University of Eichstaett-Ingolstadt [33]McMaster University

Author notes / acknowledgements

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org. The project's data is made available at the project's OSF page – see 'data availability' section for details. Other than the first three and the last author, the order of authors is alphabetical.

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 Canada Research Chair (Tier 2; PI: V. Kuperman); the Insight grant 435-2021-0657 (PI: V. Kuperman); Azrieli Early Career Faculty Fellowship (PI: N. Siegelman); NSERC Discovery Grant (PI: D. Titone); Russian Science Foundation (RSF) #21-18-00429 (PI: N. Slioussar); The Indian Institute of Technology Kanpur; The Ministry of Education, Science and Technological Development of the Republic of Serbia #451-03-9/2021-14/200163; Israel Science Foundation (ISF), #705/20 (PI: R. Frost); Faculty Research Grant, Faculty of Humanities and Social Sciences, Victoria University of Wellington, #226239 (PI: I. Elgort); PSC-CUNY Award #64464-00-52 (PI: I. A. Sekerina); Chinese Language and Technology Center of National Taiwan Normal University (NTNU) (PI: Y. T. Sung); Tomsk State University Development Programme (Priority2030); Italian Ministry of Education and Research (PRIN), #2017W8HFRX (PIs: V. Pirrelli and D. Crepaldi); and Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) #HD091013 (PI: D. Compton).

We wish to thank the following individuals: Nadine Abdelrahman, Blake Anderson, Alexander Dolge, Monica Fantini, Madison Lester, Yue Yu Liao, Chih-Tsen Liu, Yaara Loyfer, Iva Štefanija Slosar, Roni Stein, and Paul Warren.

Correspondence concerning this article should be addressed to Noam Siegelman, The Hebrew University of Jerusalem, Mount Scopus Campus, Jerusalem, Israel, 9190501, E-mail: noam.siegelman@mail.huji.ac.il

Abstract

This paper presents the ENglish Reading Online (ENRO) project, which offers data on English reading and listening comprehension from N=7,338 university-level advanced learners and native speakers of English representing 19 countries. The database also includes estimates of reading rate and seven component skills of English (including vocabulary, spelling, and grammar), as well as rich demographic and language background data. We first demonstrate high reliability for ENRO tests and their convergent validity with existing meta-analyses. We then provide a bird's-eye view of L1-L2 comparisons and examine the relative role of various predictors of reading and listening comprehension and reading speed. Across analyses, we found substantially more overlap than differences between L1 and L2 speakers, suggesting that English reading proficiency is best considered across a continuum of skill, ability, and experiences spanning L1 and L2 speakers alike. We end by providing pointers to how ENRO data can be mined for future research.

Keywords: Reading; Second language proficiency; Bilingualism; Cross-linguistic research; Open science.

1. Introduction

Bilingualism (and even multilingualism) is the norm in many countries and professions. Half the word's population speaks more than one single language (Grosjean, 2008). For example, virtually all authors of the present article are bilingual. Since English is the lingua franca of scholarly communication, it has become difficult to participate in research without a certain level of proficiency in English (Blasi et al., 2022). This is one of the reasons why knowledge of English is becoming mandatory for university students in many countries, even if they do not aspire to a research career.

Although research on acquisition and use of second language (L2) is growing both in the scope of topics and the number of publications (see bibliometric analysis in Kuperman et al., 2022), it is still fairly limited in its coverage of languages studied (both in terms of participants' L1s and the target L2s). For instance, Melby-Lervåg and Lervåg's (2014) comprehensive review of 82 studies comparing reading comprehension and other component skills of reading across L1 and L2 samples identified 11 unique L2s and 21 unique L1 backgrounds (including designations like "mixed" and "Asian"). To put this coverage in perspective, it accounts for roughly one-third of the diversity of L1 backgrounds among undergraduate students at the mid-sized Canadian McMaster University (enrolment of 30,000) and about one-half of the languages taught at that university as L2. Limited coverage aside, existing studies often use different assessments, measurement procedures, and inclusion/exclusion criteria (de Cat et al., 2022; Surrain & Luk, 2019). This variability reduces the comparability of findings across studies (de Bruin, 2019).

Another bias in existing studies is that they tend to focus on differences between L1 and L2 processing, rather than similarities, even if few researchers would sign up for the idea that L1-L2 differences are categorical rather than gradient. Such a focus is understandable within a

research enterprise strongly dominated by Popperian falsification (Brysbaert et al., 2016; Brysbaert & Rastle, 2021), where rejection of similarity (null hypothesis) is the engine of scientific progress. Null-hypothesis testing is by definition geared towards finding differences between groups and conditions rather than similarities, and requires a minimum of 860 participants to test the null effect (Cohen's $d < .2$) between two groups, and four times this number to test the absence of an interaction ($d = .2$ vs. $d = 0$) between a between-groups variable and a repeated measures variable (Brysbaert, 2019). Since such sample size is hard to achieve, practicalities of research have dictated a focus on group differences rather than similarities. Thus, Chapter 2 in an influential book on L2 reading (Grabe & Stoller, 2019, now in its third edition) contains 21 pages addressing differences between L1 and L2 reading vs. one page outlining similarities. A similar (implicit) bias emerges in meta-analyses of L2 reading or listening comprehension (see references below). By omitting L1 speakers from consideration, they can only speak to the possible differences between L1 and L2 speakers rather than their potential similarities.

Importantly, a bias towards investigating L1-L2 differences risks overlooking what is *shared* between L1 and L2 language processing, which arguably is equally important theoretically and practically. It may also lead to wrong conclusions, if hypothesis testing occurs without considering the wider context of phenomena, samples, and languages (see Brysbaert et al., 2016; Scheel et al., 2021). One example of how hypothesis-testing may over-emphasize L1-L2 differences can be found in research on cross-language neighborhood effects. Connectionist models suggest that visual recognition of a word (e.g., "beard") must overcome the competition with similarly looking words ("heard, board, bears"). Testing this hypothesis in the L2 reading domain, van Heuven et al. (1998) reported that Dutch-English bilinguals took longer to

recognize English words with many Dutch neighbors (e.g., “poor”, which has many Dutch neighbors, “boor, door, goor, hoor, koor, moor…”) than English words with few Dutch neighbors (such as “bath”, which has no Dutch words as neighbors). This was interpreted as evidence for strong inhibitory cross-language interactions in word identification (see also Whitford & Joanisse, 2021; Whitford & Titone, 2019). A large-scale follow-up study by Lemhöfer et al. (2008) compared English word recognition in native English speakers, Dutch-English bilinguals, French-English bilinguals, and German-English bilinguals, using a progressive demasking task. Contrary to van Heuven et al. (1998), the authors found many more commonalities between the groups than differences, with substantial overlap in reaction time patterns and in the set of significant predictors across the participant groups. In particular, there were virtually no influences of the bilinguals’ mother tongue on their responses to English words (as would be predicted from the cross-language neighbor inhibition effect). Lemhöfer et al. concluded that to understand English L2 word processing, it is more important to study the properties of the English language itself than possible interactions between English and the participants’ L1. (See also Diependaele et al., 2013 for a demonstration that apparent L1-L2 differences in the word frequency effect size disappear once L1-L2 differences in vocabulary size are accounted for).

This example illustrates the dangers of exclusively focusing on theory-derived differences without looking at the wider context, and of conducting hypothesis testing without exploratory, observational groundwork. An effect may be of great interest for a certain theory and at the same time of virtually no relevance to explaining overall inter-population differences. Without knowledge of the latter, it is tempting to (wrongly) generalize theoretical relevance to practical importance. One goal of this paper is to present the community of researchers with a

large data resource that facilitates the examination of reading of English both as L2 and L1 across diverse language backgrounds and samples.

1.1 Studying L1-L2 similarities and differences in a wider empirical context

A few recent high-powered studies pursued the goal of looking at the "big picture" of L2 processing (see among others Berzak et al., 2022; Cop et al., 2017; Kuperman et al., 2022). In contrast to earlier studies that focused on targeted L1-L2 experimental manipulations, these recent studies adopted a "mega-study" approach, collecting large-scale data from samples of L1 and L2 participants in a more natural reading task. For example, in the Multilingual Eye-Movement Corpus (MECO), Siegelman et al. (2022) and Kuperman et al (2022) measured eye movements of 543 university-level students from different L1 backgrounds reading L1 texts and English L2 texts. They found a striking dissociation between reading fluency and reading comprehension (see also Busby & Dahl, 2021). First, reading comprehension accuracy (assessed with multiple-choice questions) was very similar in L2 and L1, but reading fluency (assessed through reading rate and durational oculomotor measures) was much lower in L2 than in L1. Thus, group similarities and differences vary even across facets of the same task of reading for comprehension. Second, at the individual level, oculomotor measures of fluency in L2 English showed a very strong correlation with oculomotor measures in L1 and a very limited influence of English component skills (for the definition of component skills, see section below). So, L2 participants used the same oculomotor strategies in L2 reading as in L1 reading, even though they were reading with more effort in L2 than in L1. This suggests a greater degree of behavioral constancy within a reader exposed to different languages than prior literature suggested (see

reviews by Godfroid, 2021; Rayner, 1998). Conversely, reading comprehension in English was more strongly related to L2 component skills rather than to reading comprehension in L1.

The results of Kuperman et al. (2022) are intriguing, but also limited because the eye-tracking approach requires access to expensive special equipment, which is not available in all countries and laboratories. As a result, the number of test sites and participants is limited, which constrains the effect sizes that can be detected as well as the generalizability of the findings (see, e.g., Brysbaert, 2019; Vermeiren et al., 2022; Schönbrodt & Perugini, 2013). The present study, called ENglish Reading Online (ENRO), is administered fully online and does not involve eye-tracking, which opens up the study to many more diverse groups and bigger sample sizes.

The remainder of the Introduction has the following structure. We briefly overview the literature on known predictors of reading comprehension in L1 and L2, which motivated the selection of skill tests and assessments used in the ENRO study. We then proceed to presenting the main questions of the study and the analytical approaches taken to pursue those questions.

1.2 Predictors of reading comprehension

Reading researchers generally agree that reading comprehension can be explained by the combination of word reading (mapping visual perceptual input onto linguistic representations) and comprehension skills, commonly operationalised as listening comprehension (Jeon & Yamashita, 2014; Kim, 2017; Verhoeven & Perfetti, 2017). Models that include these two latent variables (e.g., the *Simple View of Reading;* Gough & Tunmer, 1986; Hoover & Gough, 1990), explain a large proportion of variance in reading comprehension (although reading and listening comprehension further involve several skills and knowledge components; Foorman et al., 2018; Kim, 2017; Peng et al., 2019).

Word reading (also referred to as "decoding") is a latent variable that involves efficient and accurate recognition of the orthographic form, which leads to automatic activation of its corresponding phonological and semantic representations. The other latent variable, comprehension, is associated with comprehension of oral language and foundational knowledge of vocabulary and grammar. Comprehension covers a range of lower- and higher-order processes that are largely the same in spoken and written discourse. These include accessing contextually appropriate word meanings, parsing sentences into chunks, encoding semantic propositions, and progressively building meaning from successive sentences (Graesser, et al., 1997; Perfetti & Stafura, 2014). Higher-order processes of comprehension rely on the success of lower-order processes which, in turn, depend on the quality of lexical and morphosyntactic representations. Consequently, poor knowledge of vocabulary and grammar can be a bottleneck in reading (and listening) comprehension (e.g., Droop & Verhoeven, 2003; Perfetti et al., 2005; Perfetti & Stafura, 2014; Raudszus et al., 2021). Thus, the ENRO battery includes assessments of both latent constructs, i.e., word reading and comprehension.

Importantly, both these constructs and the skills they represent have at least two inter-related facets: the quality of required knowledge and the ease of access to this knowledge (often referred to as fluency or automaticity; DeKeyser, 2020; Segalowitz, 2010). For instance, the knowledge of vocabulary and grammar pertain to quality of reading comprehension, whereas temporal measures (e.g., eye movements during reading, speed of word recognition or text reading) tap into reading fluency. Quality and fluency are not independent (e.g., Perfetti, 2007). At lower language proficiencies, reading for comprehension is less fluent, because readers' quality of foundational linguistic knowledge is suboptimal and, thus, access to this knowledge engages controlled rather than automatic processes. This more resource-intensive controlled

lower-order language processing consumes memory resources needed for higher-order processing involved in reading comprehension (due to the limited working memory capacity, e.g., Baddeley, 2012). This fluency deficit may prevent readers from engaging in higher-order processes (e.g., inference-making, building a coherent representation, evaluating a text's truth value), thus negatively affecting reading comprehension.

Research into L2 has identified several components both in the quality of knowledge and fluency domains that underpin reading comprehension. A relevant meta-analysis of the predictors of L2 reading comprehension in English and other languages was recently published by Jeon and Yamashita, 2022 (an update of Jeon & Yamashita, 2014). This meta-analysis showed strong positive correlations of L2 reading comprehension with – in decreasing order – L2 listening comprehension (r = 0.81), knowledge of L2 vocabulary (r = 0.72) and grammar (r = 0.70), L2 oral reading fluency (r = 0.64), knowledge of L2 morphology (r = 0.64), L2 phonological awareness (r = 0.61), knowledge of L2 orthography (r = 0.59), and L1 reading comprehension (r = 0.48). General cognitive resources (e.g., working memory and metacognition) only showed medium correlations with L2 reading comprehension (both around r = 0.33). Overall, these results suggest that L2 reading comprehension is more strongly correlated with L2 knowledge and skills than with L1 reading comprehension, leading the authors to conclude that "L2 reading comprehension is essentially determined by L2 language ability" (Jeon & Yamashita, 2014, p. 189). Clearly, these meta-analyses are an incredibly rich source of information, based on decades of rigorous research. Yet, their results are limited in important ways, namely that: (1) the estimates are based on different studies, often using different protocols and measures, (2) there is no information on how the different predictors correlate with one another, and (3) there is no comparison of L1 and L2 participants.

ENRO was set up to address some of these limitations, through a series of comparisons between large groups of L1 and L2 readers of English. Guided by previous findings, we used a battery of tasks (described below) to estimate several key component knowledge types and processing skills associated with L1 and L2 reading and listening comprehension. At the same time, we were unable to include all variables explored in previous research. In particular, we chose to focus primarily on language and reading measures of the studied language (English), leaving out measures of general cognitive abilities (e.g., working memory) and L1 performance for L2 speakers of English (e.g., L1 reading comprehension). We also omitted some measures due to time considerations and constraints of online data collection (e.g., morphological and phonological awareness). Despite these omissions, the resulting battery provides information on reading comprehension, listening comprehension, reading rate, and multiple key component skills of English proficiency, as discussed below.

1.3 The ENRO study

The ENRO study seeks to address outstanding questions of L2 reading research by offering a data resource of an unprecedented scope of language backgrounds and samples, to serve as a testbed for both hypothesis-building through data exploration and hypothesis-testing. The goals of this paper are two-fold: to introduce ENRO to reading researchers and to investigate similarities and differences in reading patterns in L1 and L2.

The first goal is achieved through open-access publication of the full data and a series of basic analyses, including reliability reports of ENRO measures and descriptive statistics broken down by participants' site, language spoken at the educational institute, and L1-L2 status.

A single paper can only shed partial light on the second goal. In this paper, we confined ourselves to two outstanding questions, motivated by the literature review above: (1) whether and on which skills L1 and L2 speakers differ on average, and (2) whether and in what ways the inter-relations between English skills vary between L1 and L2 readers. We present three sets of analyses to shed light on these questions. First, we use regression models to estimate how much variance is explained by the L1-L2 distinction in the various ENRO measures. Second, we apply correlational and factor analyses to L1 and L2 speakers of English and compare how similar or divergent the inter-relations are between various measures and constructs representing different facets of English proficiency in these two populations. Third, a partitioning-of-variance analysis asks to what extent skills and inter-sample differences predict reading comprehension and rate, as well as listening comprehension. This analysis quantifies the relative role played by the L1 vs. L2 contrast in reading and listening outcomes compared to a variety of cognitive, linguistic, and demographic factors. Following previous studies (e.g., Busby & Dahl, 2021; Dirix et al., 2020; Kuperman et al., 2022), we also probe into possible differences in the degree of L1-L2 overlap between reading comprehension (representing quality of knowledge) and rate (representing fluency), and between reading and listening comprehension.

All assessments and questionnaires of the ENRO database were administered in English and in an online format. Because the online administration did away with the demands of specialized equipment, we were able to reach 30 samples of participants including a total of 7,338 individuals representing 19 countries and 16 languages. These samples include a large number of both L1 and L2 readers of English (3853 and 3485, respectively). The languages of instruction in the institutes where ENRO participants are enrolled demonstrate a substantial typological diversity of oral languages and writing systems, including Chinese (Mandarin),

Japanese, Mongolian, Thai, Hindi, as well as Semitic (Arabic, Hebrew), Slavic (Russian, Serbian, Slovenian), Romance (French, Italian, Spanish), and Germanic (Dutch, German) languages.

Another advantage of ENRO is that it includes multiple samples from the same country or language. For several countries, ENRO comprises multiple samples with the same L1 language background (e.g., three samples from German and two from Italian universities), as well as samples representing the same country and university but a different status of English (e.g., L1 and L2 samples in North American universities). This variability enables researchers to tell apart influences of a specific L1 background from other sample-specific characteristics (e.g., regional variation or requirements of a specific educational institution).

The core of the ENRO database is the reading comprehension test, which measures both comprehension accuracy and reading rate (words per minute). An additional key component of ENRO is a listening comprehension test. ENRO also includes data from an additional set of seven assessments of component skills of English reading, including tests of vocabulary, spelling, orthographic and grammar knowledge, and lexical decision. ENRO further includes an extensive questionnaire of language background and experience.

In the Methods and Results sections that follow, we describe ENRO and the analyses we ran to validate the collected data and to investigate the theoretical questions reviewed above. In the General Discussion, we review the new insights provided by ENRO and elaborate on some of the further uses that can be made of the database.

2. Methods

2.1 Participants

The ENRO data include a total of N=7338 participants. These participants were recruited in 30 partner sites which contributed to the ENRO database. Of these, 28 samples were recruited via university-based laboratories and two were collected using the online crowdsourcing platform Prolific (prolific.ac, see below). The samples represented 19 countries and 16 unique "source languages", defined as the language of instruction in the university where data was collected (which occasionally diverged from the language of the country or region[1]). A total of nine of the samples had English as the source language (4 in Canada, 1 in New Zealand, 1 in the UK, 2 in the USA, and 1 sample of L1 English-speaking participants from the US, UK and Canada recruited through Prolific). While located in English-speaking countries, these sites included both L1 and L2 speakers of English (see below).

The remaining 21 samples were recruited in countries and academic institutions where English is not an official or dominant language: these include 20 university-based samples and one sample of Dutch speakers recruited via Prolific. In these 21 samples, partner sites were asked to exclude to the extent possible individuals who have had an uncharacteristically intensive exposure to English (students who lived in an English-speaking country for more than 6 months or speakers of English as family language). The rationale behind this exclusion was to avoid an artificial inflation of English L2 proficiency as found among typical university students speaking a given L1. In all samples, undergraduate students constituted the vast majority of participants. This was achieved by the preferential recruitment of such students among university-based partners and the use of respective screening filters in the crowdsourcing samples. Thus, samples were fairly homogeneous in terms of educational status and age.

Participating sites were requested to collect at least 100 participants. 26 out of 30 participating sites reached this sample size. We decided to include in the ENRO data the four

remaining sites also (N>50 in all samples) to avoid data loss. All participants included in ENRO have accuracy data from the central reading comprehension task. There are occasional missing values in other tasks, either due to technical errors (e.g., failed internet connection, server errors), and, in the case of some tasks, specific task requirements or outlier removal procedures. In Appendix S1 we list the number of participants with valid data from each measure and provide more details about reasons for missing data. As shown there, the number of missing values was generally small, with one exception: The listening comprehension task, which did not record responses between 2021-02-12 and 2021-04-22 due to a technical error of the web server. The data loss in this specific test did not affect most of the samples but affected many or even all participants in some samples (e.g., University of Ljubljana, Slovenia). Out of the total number of 7338 participants reported in this paper, 4875 participants completed all tests including listening comprehension (L1 = 2615; L2 = 2260).

Our analyses below range from the individual level, in which a single participant is the unit of analysis, to the group level. For group-level analyses, the relevant grouping criterion was an intersection of the recruitment site (specific university or crowdsourcing sample), source language, and the status of English as L1 or L2. We defined speakers of English as a first language (L1) as those who indicated that they first started to learn to speak English before the age of 5 in the language background survey, i.e., the age around which formal schooling begins in many participating countries. Below we discuss implications of this definition for the present results. For convenience, all other participants were labeled as L2 English speakers, even though for some English may be their third or fourth language. We considered separately subsamples of L1 and L2 speakers at sites that showed a substantial representation of both types of speakers. All such samples were found in Canada and the US universities, and hence each of these sites

includes sub-groups of both L1 and L2 speakers (see Table 1). A small percentage of participants (<5%) did not self-designate as L1 in the UK and New Zealand university samples: We assigned the majority status of L1 to all such participants by way of imputation. Several samples with a source language other than English contained participants who self-reported as L1 speakers of English (under our definition; i.e., learned to speak English before the age of 5), see Table 1. We imputed the majority L2 status to such participants instead of removing them from consideration, to avoid data loss. For participants who did not specify their age or reading/speech acquisition in English (see Appendix S1), we applied a similar imputation approach: If they are at a university with English as a source language, they are assigned to an L1 group; otherwise to an L2 group. In the data release on the project OSF's page (see *Data availability*, below), we include information regarding each participant's both original and imputed L1-L2 status, yet for the purpose of analyses below we use the imputed L1-L2 status throughout. We invite researchers to try out alternative imputations or analyses, enabled by open access to the full data. Also, in this paper we do not distinguish between L2 speakers of English enrolled as students in English-dominant countries/institutions (ESL) and L2 speakers of English in non-English-dominant sites (EFL); this is left for future research as well (see Limitations, below).

As noted above, each resulting sample of participants was defined as a combination of the recruitment site, the source language, and the imputed L1-L2 status of English: We refer to these groups of participants as "units" and use labels that encode the relevant information about each respective sample (see Table 1). For example, the unit label *ca_mcg_english_l1* refers to English-speaking (i.e., L1) subjects from the Canadian (ca) McGill University (mcg), where English is the dominant or primary language of instruction; while *de_du_german_l2* refers to

participants from Germany (de), Düsseldorf University (du), where German is the dominant language (in this case, all participants are imputed and labeled as L2 speakers of English).

As a summary of the ENRO sample, Table 1 lists the country and institution where the data were collected, the source language, each unit's sample size (including and excluding listening comprehension), and the number of L1 and L2 participants before and after the imputation of the English status. Additional information – mean age, gender and education breakdown, details regarding compensation, and participants' self-reported speaking and reading proficiency in English – is available in Appendix S2.

Table 1. Information regarding participants in the sample units.

	Unit	Country	University	English Status	Source Language	N	N L1	N LISN
1	ca_mcg_english_L1	Canada	McGill U	L1	English	61	54	60
2	ca_mcg_english_L2	Canada	McGill U	L2	English	51	0	47
3	ca_mcm_english_L1	Canada	McMaster U	L1	English	1895	1752	1743
4	ca_mcm_english_L2	Canada	McMaster U	L2	English	303	0	280
5	ca_ua_english_L1	Canada	U of Alberta	L1	English	271	246	263
6	ca_ua_english_L2	Canada	U of Alberta	L2	English	99	0	99
7	ca_uo_english_L1	Canada	U of Ottawa	L1	English	759	670	661
8	ca_uo_english_L2	Canada	U of Ottawa	L2	English	193	0	172
9	crowd_english_L1	USA, UK, CA	mixed	L1	English	299	288	290
10	nz_uv_english_L1	New Zealand	Victoria U of Wellington	L1	English	120	119	120
11	uk_uos_english_L1	UK	U of Southampton	L1	English	122	114	112
12	usa_csi_english_L1	USA	CUNY	L1	English	179	155	138
13	usa_csi_english_L2	USA	CUNY	L2	English	31	0	25
14	usa_msu_english_L1	USA	Michigan State U	L1	English	147	141	141
15	usa_msu_english_L2	USA	Michigan State U	L2	English	25	0	24
16	ar_utdt_spanish_L2	Argentina	U Torcuato Di Tella	L2	Spanish	102	28	94
17	be_ugh_dutch_L2	Belgium	U Ghent	L2	Dutch	205	7	199
18	be_ulb_french_L2	Belgium	U Libre de Bruxelles	L2	French	105	5	68
19	be_nl_crowd_dutch_L2	Belgium, NL	mixed	L2	Dutch	193	24	191
20	de_du_german_L2	Germany	Heinrich-Heine-U Düsseldorf	L2	German	53	5	51
21	de_gu_german_L2	Germany	U Goettingen	L2	German	146	7	145
22	de_ku_german_L2	Germany	Katholische U Eichstatt-Ingolstadt	L2	German	104	4	102
23	il_huji_he_hebrew_L2	Israel	Hebrew U	L2	Hebrew	112	8	111
24	il_huji_ar_arabic_L2	Israel	Hebrew U	L2	Arabic	101	12	100
25	in_iitk_hindi_L2	India	Indian Institute of Tech, Kanpur	L2	Hindi	157	70	157
26	it_si_italian_L2	Italy	SISSA	L2	Italian	151	7	148
27	it_unimib_italian_L2	Italy	U of Milano-Bicocca	L2	Italian	221	34	204
28	jp_nu_japanese_L2	Japan	Nagoya U	L2	Japanese	129	8	128
29	mn_kho_mongolian_L2	Mongolia	Khovd State U	L2	Mongolian	51	15	51
30	ru_hse_russian_L2	Russia	HSE Moscow	L2	Russian	73	8	61

	Unit	Country	University	English Status	Source Language	N	N L1	N LISN
31	ru_spb_russian_L2	Russia	St Petersburg U	L2	Russian	59	6	59
32	ru_tu_russian_L2	Russia	Tomsk U	L2	Russian	164	8	164
33	rs_bg_serbian_L2	Serbia	U of Belgrade	L2	Serbian	301	45	296
34	si_lj_slovene_L2	Slovenia	U of Ljubljana	L2	Slovene	102	9	100
35	th_tu_thai_L2	Thailand	Thammasat U	L2	Thai	101	35	99
36	tw_ntnu_chinese_L2	Taiwan	National Taiwan Normal U	L2	Chinese	153	23	152

Notes: N L1- Number of participants in the unit who reported learning English before the age of 5; N LISN- Number of participants with listening comprehension data; U-University; USA-United States of America; UK-United Kingdom; CA-Canada; NL-Netherlands; CUNY- City University of New York College of Staten Island; M/F/O- Male/Female/Other.

The project-wide ethics clearance was obtained through the Research Ethics Board of McMaster University (protocol #4968). Each individual partner site additionally obtained an ethics clearance or a waiver from the ethics research board of the corresponding institution or country. We only include data from participants who did not withdraw in the course of the study and allowed the use of their (de-identified) data.

2.2 Materials

All ENRO participants completed the same battery of instruments in English including tests of reading and listening comprehension, tests of multiple component skills of reading, a motivation survey, and demographic and language background questionnaires. Below we describe the major ENRO instruments (reading comprehension and listening comprehension tasks; demographic and language background questionnaire). Description of additional ENRO instruments (tests of component skills and motivation survey) are available in Appendix S3. Estimates of ENRO measures' reliability follow in the Results section.

2.2.1 Reading comprehension

Participants read a set of 15 texts in English. Texts were based on training materials for the ACCUPLACER Reading test and the English as Second Language Reading Skills test (https://accuplacer.collegeboard.org/students/prepare-for-accuplacer/practice), which are commonly used for course placement of L1 and L2 speakers of English in North American colleges. All texts were written in expository prose and presented information about a person (e.g., Samuel Morse) or a historic or natural phenomenon (e.g., Da Vinci's inventions). Each text was followed by three 4-alternative-forced-choice comprehension questions designed to test

individuals' ability to determine central ideas of the text, summarize and synthesize its content and analyze argumentation, word choice and text structure. Appendix S4 further provides details regarding number of sentences and words in each text, as well as their estimated readability measures.

Texts and questions were presented to participants in a fixed order. The measure of *reading comprehension* was the participant's percent of correct responses out of the 45 questions. Twelve out of the 15 texts in the reading comprehension texts were also used in the L2 component of the Multilingual Eye-Movement Corpus (labeled MECO-L2, Kuperman et al., 2022). Thus, the ENRO reading comprehension accuracy data can be used to produce measures that are backward compatible with MECO L2 comprehension scores: The project's OSF page contains these scores for those interested in direct ENRO-MECO comparisons.

Additionally, we measured *reading rate* as the number of words in each text divided by its total reading time (words per minute, wpm). Values that were unrealistically long (possibly reflecting distraction or connectivity issues) or short (possibly reflecting a response after only partial reading, skimming, or skipping) were disregarded. We only considered reading rates in the interval between 89 and 804 wpm, i.e., reading rates that are 3 times slower and faster than the estimated average reading rate of an L1 reader (268 wpm; see Brysbaert, 2019 and Just and Carpenter, 1987; see also Kuperman et al., 2021). We then computed a measure of average reading rate across texts for each subject.

2.2.2 Listening comprehension

The test was an adaptation of the Lectures, Interviews and Spoken Narratives (LISN) listening comprehension test (Sommers et al., 2011; Tye-Murray et al., 2008), which consisted of 5 recorded audio passages based on the narratives selected from the Rutgers University Oral

History Archives of personal descriptions of life experiences (Sommers et al., 2011). The passages were between 1–2 minutes long and were recorded by male and female professional actors with North American accents. ENRO uses edited versions of 5 out of the original 16 narrative passages. Appendix S4 includes information regarding the passages' readability estimates. Each text was followed by five 4-alternative-forced-choice comprehension questions: defined by Sommers et al. (2011) as information questions (asking to recall a specific factual piece of information from the passage), integration questions (designed to assess ability to combine two or more separate pieces of textual information), and inference questions. The percent of questions answered correctly out of the 25 questions was the participant's score in the task (labeled below as *listening comprehension*).

2.2.3 Demographic and Language Background Questionnaire

All participants completed a Brief Social and Language History Questionnaire (B-LSHQ), aimed at collecting basic demographic and linguistic information for both their L1 and (in case of English L2 speakers) English as their L2. The questionnaire, designed by the McGill Language and Multilingualism laboratory, corresponds to an abridged version of the questionnaire reported in Gullifer and Titone (2020), which was originally adapted from various questionnaires used in the field of bilingualism (in particular, the Language Experience and Proficiency Questionnaire, LEAP-Q; Kaushanskaya et al., 2020; and the Language History Questionnaire 3.0, LHQ-3.0, Li et al., 2020).

The first part of the survey included questions about participants' age, gender, university, degree, year of study and years of education. The second part of the survey included two sections with questions about the languages each participant reads and speaks. In each of these sections,

the participant first listed the languages they speak/read (either by selecting them from a dropdown menu of languages, or by typing them manually). Then, for languages the participant reported speaking and reading, questions assessed language-usage patterns for a range of communicative contexts (at home, at school, at work, in public, with family, with friends, and when applicable, with a significant other). In line with Gullifer and Titone (2021a), percent-based scales were used (e.g., "Indicate your current percentage use of all of the languages you speak, in each of the following environments").

2.2.4 Component skills of English reading proficiency and a motivation questionnaire

Seven tests tapped into component skills of English reading proficiency: Grammaticality judgment task, spelling recognition, vocabulary knowledge, orthographic awareness, text segmentation, the Lexical Test for Advanced Learners of English (LexTALE), and lexical decision task. The eighth test tapped into how motivated our participants were to excel in the study. As noted above, the tests were chosen because of their theoretical relevance (i.e., to provide measures of word reading and listening comprehension; as well as foundational language knowledge), and in view of practical considerations of an online administration and study duration. Several of these component skills are equally relevant for listening comprehension (e.g., vocabulary knowledge, grammatical knowledge), while others are specific to the written modality (e.g., spelling, orthographic awareness, text segmentation). In some cases, multiple tests tapped into the same or related latent constructs (e.g., LexTALE and Vocabulary Knowledge), and can therefore be used to gauge ENRO's convergent validity. Other considerations behind test selection included maintaining backward compatibility with MECO-L2 (which had the same measures of vocabulary knowledge, spelling recognition, and

LexTALE). Details on the tests' stimuli and scoring, including references, are provided in Appendix S3.

2.3 Procedure

The study was administered online, using an in-house web-based data collection platform. Participants began with reading the project-wide standard consent form in English; some partner sites added a second consent form, as required by their local ethics research board. Participants then proceeded to the test battery. At any point, they could withdraw from the study. The tasks in the battery were presented in the following fixed order: (1) demographic and language background questionnaire; (2) reading comprehension; (3) grammaticality judgment task; (4) listening comprehension; (5) spelling recognition; (6) vocabulary knowledge; (7) motivation, (8) orthographic awareness; (9) text segmentation; (10) LexTALE; and (11) lexical decision. The entire study typically took about 1.5 hours to complete, rarely exceeding 2 hours.

2.4 Data availability

The ENRO project is committed to principles of Open Science. The current release includes the full data from all participants on all tests and questionnaires. Reports are available both at the participant-level (i.e., each participant's performance by test) and, in applicable tests, at the trial-level, and the analytical code is provided as well. Please refer to the project's repository page at https://osf.io/gzyqf/ for the full code and data.

3. Results

This section is organized into five sections. Sections 3.1 and 3.2 provide the methodological foundations for current and prospective use of ENRO data: Section 3.1 estimates the reliability of ENRO tests, and section 3.2 provides descriptive statistics per participant "unit". Then, in sections 3.3-3.5 we proceed to analyses that tackle the main theoretical questions motivated in the Introduction. Section 3.3 quantifies how much of the variance in each task is explained by overall group differences between L1 and L2 readers of English. Thus, it quantifies whether and to what extent L1 and L2 readers differ, on average, in the various ENRO tests. Section 3.4 focuses on the inter-relations between ENRO tasks and the underlying latent constructs they tap into, and examines whether and how they are different in L1 and L2 participants. Finally, section 3.5 examines the relative role of different variables in predicting three major outcomes of interest, i.e., English reading comprehension, listening comprehension, and reading rate. It thus quantifies the relative contributions of component skills, L1-L2 differences, and additional characteristics of a participant's background as predictors of these outcomes.

3.1 Reliability Estimates

To confirm that ENRO tests are sensitive enough for individual-differences analyses, we computed reliability estimates for each of the ENRO measures.[2] We used a split-half procedure in all tests, with the exception of reading rate where the data structure was more fitting of an Intraclass Coefficient (ICC) analysis (i.e., examining agreement in reading rates across a relatively small number of 15 texts). For split-half estimates, we report Spearman-Brown corrected values (Spearman, 1910), which reflect reliability for the full sample of items in the test (rather than for half of the items, which are the bases for uncorrected correlation estimates).

Table 2 shows that reliability was high for all ENRO measures in the full sample of participants (all estimates ≥ 0.74), and very high for main variables of interest (reading comprehension, reading rate, and listening comprehension; all estimates ≥ 0.80). Table 2 also provides reliability estimates computed separately for L1 and L2 participants (based on our measure of "imputed" L1-L2 status). These estimates confirmed that the measures' sensitivity is of satisfactory levels for both L1 and L2 participants (all estimates ≥ 0.69).

Table 2. Reliability estimates for ENRO measures. For all measures but reading rate a split-half procedure was used; estimates in the table are Spearman-Brown's corrected. For reading rate we used ICC across the 15 texts. Estimates computed separately for L1 and L2 participants are also provided.

Measure	Reliability: full sample	Reliability: L1	Reliability: L2
Reading Comprehension	0.85	0.86	0.85
Reading Rate (Comprehension task texts)	0.96	0.95	0.96
Listening Comprehension	0.80	0.80	0.80
Motivation	0.81	0.83	0.78
Vocabulary	0.93	0.94	0.92
Spelling	0.81	0.80	0.80
Grammatical Knowledge	0.81	0.69	0.79
Lexical Decision: Accuracy	0.98	0.98	0.96
Lexical Decision: RT	0.98	0.98	0.97
LexTALE: Accuracy	0.90	0.90	0.86
LexTALE: RT	0.95	0.91	0.96
Orthographic Awareness	0.74	0.76	0.72

3.2 Descriptive Statistics: Mean Reading Performance by Participant Unit

This section provides descriptive statistics across *units* of participants (i.e., samples of participants defined by their recruitment site, source language, and imputed status of English as L1 or L2, see above). To this end, we calculated the means and standard errors for all ENRO measures for each unit. For brevity, in the main text we provide descriptive plots for two central outcomes of interest: Reading comprehension (Figure 1) and reading rate (Figure 2). Figures with descriptive plots for all other ENRO measures are available in Appendix S5. Also, we make available via the project's OSF page an auxiliary table with full descriptive information broken down by unit (e.g., number of participants, mean, standard deviation, median, range, and standard error for each measure).

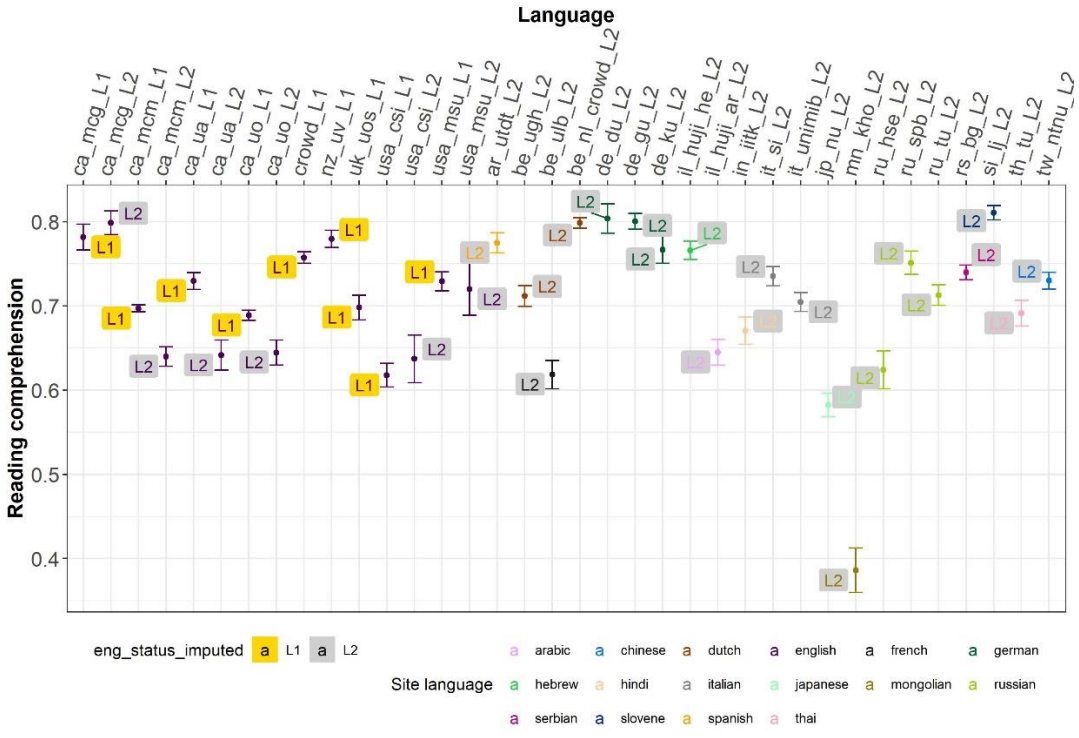

Figure 1. Reading comprehension scores, proportion correct. Error bars stand for ±1 SE.

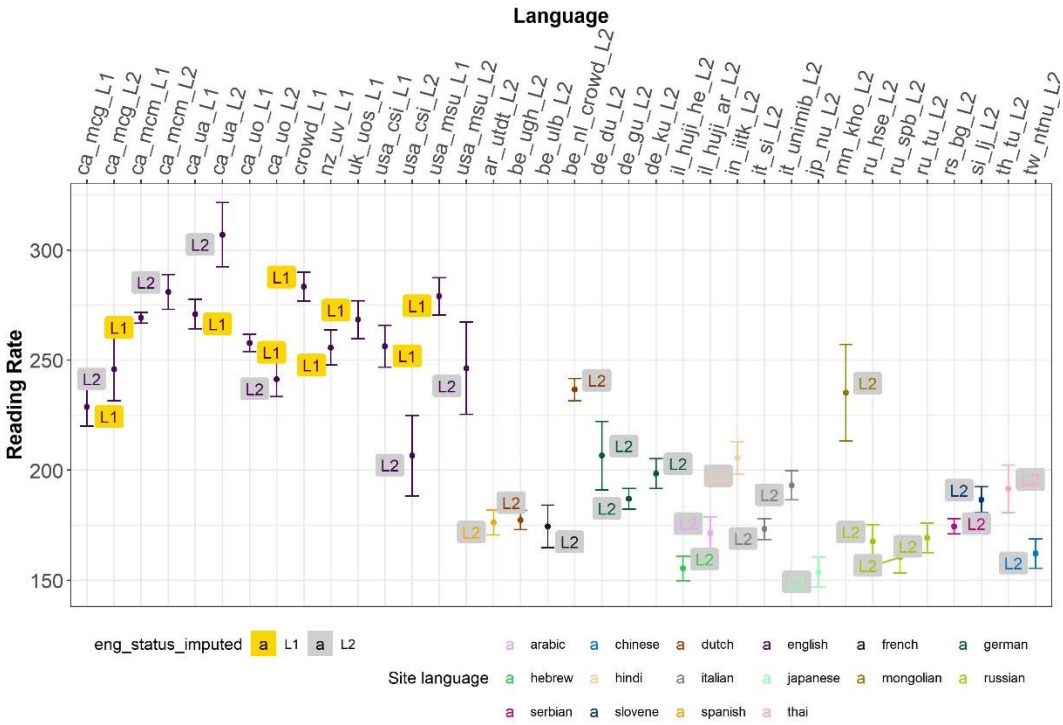

Figure 2. Reading rate, words per minute (wpm). Error bars stand for ±1 SE.

A few noteworthy trends emerge from Figures 1 and 2. The difference in English reading comprehension accuracy between English-dominant (universities in Canada, New Zealand, UK, and US) and non-dominant sites appears to be small. This replicates the finding reported by Kuperman et al. (2022). Yet, reading comprehension varied considerably between testing sites representing the same languages (whether one considered L1 or L2 speakers of English). It is worth noting that similar patterns were observed in listening comprehension performance (see Appendix S5). In contrast, just like in MECO-L2 (Kuperman et al., 2022), estimates of mean reading rate (Figure 2) demonstrate a clear separation in performance in English-language tasks between the sites where English is and is not a dominant language, with noticeably faster reading (more words per minute) in English-dominant sites.

3.3 Comparison of L1 and L2 readers' mean performance across tasks

This analysis asks how much variance the L1-L2 distinction explains in participants' performance in all experimental tasks of ENRO tapping into the English language processing. This information is useful for understanding which component skills or outcomes of reading show greater or lesser overlap between these groups of participants. Table 3 shows the means and standard deviations of L1 and L2 readers on all tasks, together with the standardized effect size (Cohen's *d*) of that difference and the percentage of variance explained by (imputed) language status (L1-L2) in an ordinary regression model fitted to the respective test score. In this analysis, we only included participants without missing values in any of the dependent variables, to ensure that estimates are based on the same set of participants for all measures (N=5023).

Table 3. Means and SDs of L1 and L2 readers**,** standardized effect size of the L1-L2 difference, and percent variance explained by the imputed L1-L2 status, for all measures of English proficiency.

Measure	L1 participants: Mean (SD)	L2 participants: Mean (SD)	Cohen's *d*	Percent variance explained
Grammatical Knowledge	0.87 (0.09)	0.74 (0.15)	1.12	24%
LexTALE: Accuracy	0.86 (0.11)	0.74 (0.13)	1.01	20%
Vocabulary	62.25 (9.11)	51.63 (13.74)	0.92	17%
Lexical Decision: Accuracy	0.85 (0.12)	0.75 (0.12)	0.79	14%
Reading Rate (Comprehension task texts)	267.3 (106.35)	198.54 (94.31)	0.68	10%
Text Segmentation	41.31 (13.14)	32.92 (13.58)	0.63	9%
LexTALE: RT	968.96 (283.35)	1200.55 (475.23)	0.6	8%
Lexical Decision: RT	720.63 (114.43)	802.76 (158.63)	0.6	8%
Spelling	0.86 (0.09)	0.81 (0.11)	0.48	5%
Listening Comprehension	0.64 (0.18)	0.6 (0.2)	0.22	1%
Orthographic Awareness	0.89 (0.1)	0.87 (0.09)	0.2	1%
Reading Comprehension	0.73 (0.15)	0.73 (0.15)	0	<1%
Motivation	3.59 (0.6)	3.57 (0.55)	-0.03	<1%

This analysis quantifies the "main effect" of the L1-L2 distinction on all English-language skill tests collected in ENRO. Component skills of reading that stood out as particularly impacted by L1-L2 differences were grammatical knowledge of English (24% variance explained), LexTALE accuracy (20%), and vocabulary knowledge (17%). On the other extreme, L1 and L2 speakers of English showed the same mean motivation to excel (<1%), orthographic awareness (1%), and listening comprehension (1%) scores. Spelling performance was another component skill that only weakly differentiated whether a participant was a native speaker of English or not (5%). Confirming the visual inspection of Figures 1 and 2, the L1-L2 status explained little to no variance in reading comprehension (<1%) but was a strong predictor of reading rate (10%). The resulting hierarchy of effects indicates where the coarse-grained differences between groups of L1 and L2 readers lie and paves the way for more in-depth analyses.

3.4 Inter-relations between English skills among L1 and L2 readers

We next turn to analyses quantifying the inter-relations between ENRO measures, and, primarily, the extent to which these inter-relations vary between L1 and L2 readers. Thus, complementary to the estimates of "main effects" in the last section (i.e., the overall differences in mean performance between L1 and L2 readers), analyses in this section tap into interactions of the L1-L2 status and English component skills as predictors of English reading comprehension and rate, and English proficiency more broadly.

Correlational analysis: Table 4 provides a matrix with all pairwise Pearson's correlations between test scores, computed separately for L1 and L2 readers. Correlations between scores among L1 readers and L2 readers are shown above and below the diagonal, respectively. In Appendix S6 we further provide estimates for the correlations after correction for attenuation given the tests' reliability, as well as correlation estimates computed over the full sample (i.e., including both L1 and L2 readers).

Table 4. Correlations between ENRO measures. Values above the diagonal show correlations among L1 participants (maximum N=3853) and below the diagonal among L2 participants (maximum N= 3485). All values are uncorrected Pearson's correlations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) read_comp_score		0.32	0.69	0.65	0.45	0.46	0.55	0.27	0.62	0.31	0.42	0.47	-0.3	0.11	0.1	-0.08	-0.1
(2) motivation	0.34		0.36	0.23	0.13	0.16	0.23	0.14	0.27	0.14	0.15	0.18	-0.13	0.03	0.04	-0.07	-0.04
(3) lisn_score	0.68	0.33		0.53	0.38	0.4	0.51	0.26	0.54	0.27	0.33	0.39	-0.22	0.07	0.07	-0.05	-0.08
(4) vocab_score	0.59	0.29	0.55		0.49	0.53	0.59	0.22	0.63	0.29	0.45	0.48	-0.11	0.2	0.18	-0.16	-0.15
(5) spell_score	0.48	0.23	0.42	0.61		0.36	0.45	0.1	0.47	0.14	0.36	0.4	-0.01	0.19	0.22	-0.06	-0.09
(6) grammar_score	0.49	0.24	0.61	0.53	0.38		0.47	0.08	0.46	0.13	0.34	0.4	-0.12	0.22	0.19	-0.15	-0.14
(7) lextale_score	0.52	0.27	0.56	0.62	0.53	0.59		0.21	0.77	0.26	0.42	0.46	-0.1	0.19	0.17	-0.15	-0.13
(8) lextale_rt	0.21	0.08	0.13	0.13	0.12	-0.05	0.13		0.27	0.63	0.18	-0.03	-0.31	-0.03	-0.06	0	-0.02
(9) ld_score	0.61	0.29	0.62	0.63	0.55	0.63	0.77	0.13		0.44	0.45	0.46	-0.14	0.15	0.14	-0.14	-0.13
(10) ld_rt	0.32	0.14	0.2	0.22	0.19	0.07	0.18	0.64	0.38		0.21	-0.04	-0.31	-0.01	-0.01	-0.01	-0.02
(11) ortho_score	0.4	0.15	0.29	0.34	0.37	0.26	0.31	0.12	0.41	0.21		0.34	-0.1	0.09	0.09	-0.06	-0.08
(12) segment_score	0.49	0.24	0.47	0.43	0.44	0.46	0.51	-0.08	0.57	-0.01	0.35		0	0.12	0.13	-0.13	-0.15
(13) rate_mean	-0.26	-0.08	-0.17	0.03	0	-0.01	0.03	-0.21	-0.06	-0.27	-0.13	-0.01		0.05	0.08	-0.03	-0.03
(14) eng_speech_proficiency	0.27	0.19	0.4	0.41	0.32	0.55	0.45	-0.17	0.42	-0.08	0.13	0.32	0.06		0.77	-0.16	-0.11
(15) eng_read_proficiency	0.27	0.21	0.37	0.39	0.35	0.5	0.43	-0.18	0.41	-0.1	0.15	0.33	0.06	0.82		-0.1	-0.11
(16) eng_speech_age	-0.1	0	-0.09	-0.08	-0.1	-0.06	-0.09	0.06	-0.09	0.03	-0.06	-0.15	-0.01	-0.15	-0.12		0.42
(17) eng_read_age	-0.07	-0.01	-0.08	-0.04	-0.08	-0.07	-0.06	0.03	-0.06	0.02	-0.07	-0.11	-0.04	-0.13	-0.13	0.61	

Notes: read_comp_score: Reading Comprehension; lisn_score: Listening Comprehension; vocab_score: Vocabulary Knowledge; spell_score: Spelling; grammar_score: Grammatical Knowledge; lextale_score: LexTALE, accuracy; lextale_rt: LexTALE, mean RT; ld_score: Lexical Decision, accuracy; ld_rt: Lexical Decision, mean RT; ortho_score: Orthographic Awareness; segment_score: Text Segmentation; rate_mean: Reading Rate (Comprehension task texts); eng_speech_proficiency: Self-rated proficiency, English, speech; eng_read_proficiency: Self-rated proficiency, English, reading; eng_speech_age: Age of English speech onset; eng_read_age: Age of English reading onset.

Table 5. Comparison of the magnitudes of correlations among L1 and L2 participants. Values above the diagonal show correlation difference (r_{L1} *minus* r_{L2}). Values below the diagonal are *p*-values of each difference (based on a Fisher's *r*-to-*Z* transformation). *p*-values written as 0.0000 are <.0001, and difference in *r* listed as 0.00 are <|.01|. Correlation differences significant after Bonferroni correction (α=5%/136) are shown in bold.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) read_comp_score		-0.02	0.00	**0.06**	-0.03	-0.04	0.02	0.06	0.00	-0.01	0.02	-0.02	-0.03	**-0.16**	**-0.16**	0.02	-0.04
(2) motivation	0.3371		0.03	-0.06	**-0.10**	-0.08	-0.04	0.06	-0.02	0.00	-0.01	-0.06	-0.06	**-0.15**	**-0.17**	-0.07	-0.03
(3) lisn_score	0.4833	0.2052		-0.02	-0.04	**-0.21**	-0.05	**0.13**	**-0.08**	0.07	0.05	**-0.09**	-0.06	**-0.33**	**-0.30**	0.03	0.01
(4) vocab_score	0.0000	0.0061	0.2951		-0.12	0.00	-0.03	**0.09**	0.00	0.07	**0.11**	0.05	**-0.14**	**-0.21**	**-0.22**	-0.08	**-0.10**
(5) spell_score	0.1023	0.0000	0.0769	0.0000		-0.02	**-0.08**	-0.02	**-0.08**	-0.05	-0.02	-0.04	-0.01	**-0.13**	**-0.13**	0.04	-0.01
(6) grammar_score	0.0998	0.0004	0.0000	1.0000	0.3256		**-0.12**	**0.13**	**-0.17**	0.06	**0.08**	-0.06	**-0.11**	**-0.33**	**-0.32**	**-0.09**	-0.07
(7) lextale_score	0.0731	0.0689	0.0094	0.0443	0.0000	0.0000		0.09	0.00	**0.09**	**0.11**	-0.05	**-0.13**	**-0.26**	**-0.26**	-0.07	-0.07
(8) lextale_rt	0.0066	0.0096	0.0000	0.0001	0.3892	0.0000	0.0004		0.14	-0.01	0.05	0.05	**-0.10**	**0.14**	**0.11**	-0.06	-0.05
(9) ld_score	0.4933	0.3550	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000		0.07	0.04	**-0.11**	-0.08	**-0.27**	**-0.27**	-0.05	-0.07
(10) ld_rt	0.6362	1.0000	0.0060	0.0015	0.0287	0.0103	0.0003	0.4752	0.0021		0.00	-0.03	-0.04	0.07	0.09	-0.04	-0.03
(11) ortho_score	0.3063	1.0000	0.1012	0.0000	0.6241	0.0002	0.0000	0.0091	0.0372	1.0000		-0.01	0.02	-0.04	-0.07	0.01	-0.01
(12) segment_score	0.2735	0.0081	0.0003	0.0080	0.0410	0.0020	0.0060	0.0351	0.0000	0.2073	0.6334		0.01	**-0.20**	**-0.20**	0.02	-0.03
(13) rate_mean	0.0692	0.0343	0.0580	0.0000	0.6759	0.0000	0.0000	0.0000	0.0007	0.0683	0.2052	0.6796		-0.01	0.02	-0.01	0.01
(14) eng_speech_proficiency	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0047	0.1033	0.0000	0.6906		**-0.06**	-0.01	0.01
(15) eng_read_proficiency	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0004	0.0173	0.0000	0.4383	0.0000		0.01	0.02
(16) eng_speech_age	0.4151	0.0046	0.1574	0.0011	0.1044	0.0003	0.0142	0.0156	0.0414	0.1072	1.0000	0.4156	0.4272	0.6789	0.4266		**-0.20**
(17) eng_read_age	0.2356	0.2388	1.0000	0.0000	0.6928	0.0058	0.0057	0.0505	0.0057	0.1176	0.6943	0.1146	0.6995	0.4255	0.4255	0.0000	

Notes: read_comp_score: Reading Comprehension; lisn_score: Listening Comprehension; vocab_score: Vocabulary Knowledge; spell_score: Spelling; grammar_score: Grammatical Knowledge; lextale_score: LexTALE, accuracy; lextale_rt: LexTALE, mean RT; ld_score: Lexical Decision, accuracy; ld_rt: Lexical Decision, mean RT; ortho_score: Orthographic Awareness; segment_score: Text Segmentation; rate_mean: Reading Rate (Comprehension task texts); eng_speech_proficiency: Self-rated proficiency, English, speech; eng_read_proficiency: Self-rated proficiency, English, reading; eng_speech_age: Age of English speech onset; eng_read_age: Age of English reading onset.

Table 4 already points to a good convergence between the ENRO data and results of the meta-data analyses of reading and listening comprehension by Jeon and colleagues (Jeon & Yamashita, 2014; 2022; In'nami et al., 2022). Specifically, the order of the predictors of L2 reading comprehension, ranked by the strength of the correlation, was extremely similar between previous meta-analyses and our data, with L2 reading comprehension most strongly predicted by listening comprehension, followed by tests of vocabulary knowledge (also gauged through lexical decision and LexTALE tests), grammar knowledge, and orthographic awareness. To further compare previous meta-analytic estimates to the ENRO data, in Appendix S7 we present estimates of predictors of reading and listening comprehension in L2 participants from the meta-analyses by Jeon and colleagues (Jeon & Yamashita, 2014; 2022; In'nami et al. et al.,2022), alongside ENRO estimates for both L1 and L2 participants. Overall, the convergence of estimates among L2 participants gives additional credibility to both the measures overlapping with those in the meta-analyses, and the additional variables we introduced (motivation to excel, subjective speech and reading proficiency, and skill tests). Importantly, the similarity in the relative strength of correlations also span over the L1 participants in ENRO, who demonstrated a similarly strong role of grammar, vocabulary, and orthographic knowledge as predictors of reading and listening comprehension as L2 participants in ENRO and the meta-analytic data.

Next, and central to our theoretical question, we *compared* the correlations computed separately for L1 and L2 participants. Thus, for each of the 136 pairwise correlations, we computed a difference between correlation estimates in the two samples of participants and examined whether this difference significantly differed from 0 (using Fisher's *r*-to-*Z* transformation) after Bonferroni correction for multiple comparisons. The results are presented in Table 5 (with Δr's shown above the diagonal, and *p*-values below the diagonal). Notably, in

most cases, L1 and L2 readers showed similar magnitude of correlations between ENRO tests. In fact, out of 136 pairwise comparisons, only 12 exceeded an absolute value of 0.2, and out of these cases, 11 involved correlations with self-report measures (i.e., age of acquisition of English reading and speech; English self-rated proficiency). These differences are due to the lack of variability in responses among L1 speakers of English (e.g., the decisive majority of responses in this group reported 0 for age of acquisition of English and 7/7 for self-rated proficiency). The only correlation between experimental tasks that showed a L1-L2 difference larger than |0.2| was between grammatical knowledge and listening comprehension (L1: $r = 0.40$; L2: $r = 0.61$). We conclude that inter-relations between ENRO tests are generally similar in magnitude among L1 and L2 readers. Still, we note that some multiple correlation differences reached statistical significance: Even after excluding correlations with self-report measures, 25/78 (32%) were significant after applying a Bonferroni correction. Most of these statistically significant L1-L2 differences were practically small, and involved either grammatical knowledge (7 correlations) or LexTALE (6 in both accuracy and 6 in RT), suggesting that component skills of English proficiency that differentiated L1 and L2 readers the most in performance (Table 3) are also the ones that show the greatest difference in predictive power between L1 and L2 groups. More broadly, this result points again to the distinction between statistical, theoretical, and practical significance of L1-L2 comparisons, which we return to in the General Discussion.

Factor analysis: The correlational analysis taps into inter-relations between variables defined at the single-task level, and examines whether and how these correlations vary between L1 and L2 participants. A logical next step is to investigate how the variables load into latent factors, and whether this grouping is different between L1 and L2 participants. An initial answer to this question is provided here by an exploratory factor analysis, testing how the various tests

group together and how similar the solution is for L1 and L2 speakers (see also Gullifer et al., 2021).

Factor analysis using the default parameters of the R psych() package (minimal residual extraction combined with oblimin rotation; Revelle, 2015) indicated that more than half the variance was accounted for by three factors in both L1 and L2 participants (51% for L1 and 54% for L2). Table 6 presents the results of the two factor analyses conducted on the two samples.

Table 6. Exploratory factor analyses for L1 and L2 speakers: Variable loadings (absolute values higher than 0.3 are presented) and cumulative variance explained. Additional information can be accessed via the code at the project's OSF page.

	L1 participants			**L2 participants**		
	Factor 1	*Factor 2*	*Factor 3*	*Factor 1*	*Factor 2*	*Factor 3*
ld_score	0.85			0.81		
lextale_score	0.81			0.83		
vocab_score	0.47		0.33	0.72		
spell_score	0.44			0.66		
segment_score	0.45	-0.33		0.53		
ortho_score	0.33					
grammar_score	0.35			0.54		
ld_rt		0.83			0.91	
lextale_rt		0.74			0.75	
read_comp_score			0.84	0.32		0.56
lisn_score			0.62	0.36		0.57
rate_mean		-0.37	-0.37	0.35		-0.48
Cumulative variance explained	*22%*	*38%*	*51%*	*30%*	*43%*	*54%*

Notes: read_comp_score: Reading Comprehension; lisn_score: Listening Comprehension; vocab_score: Vocabulary Knowledge; spell_score: Spelling; grammar_score: Grammatical Knowledge; lextale_score: LexTALE, accuracy; lextale_rt: LexTALE, mean RT; ld_score: Lexical Decision, accuracy; ld_rt: Lexical Decision, mean RT; ortho_score: Orthographic Awareness; segment_score: Text Segmentation; rate_mean: Reading Rate (Comprehension task texts).

Among both L1 and L2 participants, multiple accuracy-based tests of English component skills group into the first factor: These include accuracy from lexical decision, spelling, vocabulary, grammar, and text segmentation (orthographic awareness was loaded onto this factor in L1 but not in L2 participants). Also reflecting similarity in L1 and L2 groups, the RT measures in lexical decision and LexTALE tasks (after logarithmic transformation) group into a second factor, which reflects response slowness. The two comprehension scores (listening and reading) load onto a third factor, a factor which correlates strongly (r = 0.64 and 0.55 for L1 and L2 speakers, respectively) with the "proficiency factor" (i.e., factor number 1). Finally, in both L1 and L2 participants, reading rate loads on factor 3: fast readers in both samples had lower comprehension scores. Together, the results of the factor analysis suggests that although there are subtle differences between L1 and L2 participants (e.g., loading of comprehension scores and reading rate on factor 1 in L2 but not L1 speakers; loading of reading rate on factor 2 in L1 but not L2 speakers; see more on these differences in the General Discussion), the similarities in how variables of English proficiency group to latent factors in L1 and L2 participants display substantial overlap.

3.5 What explains variance in reading and listening comprehension?

Analyses so far estimated the overall differences between L1 and L2 readers in their English-language performance, and the similarities and differences in inter-relations between test scores (and their loadings on latent factors) in the two participant groups. In this section, we tackle an additional related theoretical issue motivated in the Introduction: What are the sources of inter-individual variability in English reading and listening comprehension? Unlike comparative

quantification of the group effect of L1-L2 distinction presented above, this analysis enables us to examine the possible L1-L2 similarities and differences when controlling for other factors that explain variability in the English performance (specifically, English component skills). The analysis also estimates the predictive value of L1-L2 difference *relative* to other sources of variance.

In all analyses in this section, we treat as dependent variables three central outcome variables: Reading comprehension, listening comprehension, and reading rate. For each of these variables, we conducted a partitioning-of-variance analysis, where variability in the three outcome variables is decomposed to that explained by four factors: (1) English component skills, (2) whether a participant is or is not a native speaker of English (i.e., L1-L2 distinction), (3) the L1 of the participant (among L2 speakers), and (4) the inter-sample differences within a country (see definition below). Operationally, we used a series of successive regression models, where the dependent variable was one of the three outcome variables and a group of predictors was added at each step to examine the additional unique amount of variance explained by the predictors added: See Kuperman et al., (2022), for a similar approach.

Predictors were added to the models in four steps. At Step 1, we added variables that reflect performance in measures of English component skills. When reading comprehension was the dependent variable, we also added listening comprehension as a predictor at this step, in line with accounts that highlight the role of listening comprehension in predicting reading comprehension (e.g., the Simple View of Reading, Hoover & Gough, 1990). Variance explained at this step reflects the overall role of component skills, gauged via skill tests, in explaining our outcomes of English proficiency. At Step 2, we added the imputed L1-L2 status variable. The added variance in this step reflects the variance explained by the distinction between L1 and L2

speakers, while controlling for the effect of English component skills. At Step 3, we added a set of dummy-coded variables coding for participants' native language(s) other than English. Each variable in the set represented one native language spoken by participants in our sample, with native speakers of a given language coded as 1 and non-native speakers of that language coded as 0.[3] This step reflects added variance associated with differences between L2 English readers of different L1s (i.e., the effect of a participant's L1 on their L2 English performance), beyond the L1-L2 effect and the impact of component skills. At the final Step 4, we added a categorical variable reflecting differences between sites within a country (inserted as a factor to the model and implemented as a series of dummy-coded variables). Thus, for example, this variable distinguishes between the two different U.S. English-dominant sites (College of Staten Island and Michigan State University), and the three different sites of German speakers in Germany (Universities of Düsseldorf, Göttingen, and Eichstätt-Ingolstadt). The added variance at this step reflects additional factors that are expected to vary across sites within a country and language (e.g., differences in educational background, English entry requirements, socio-economic status, etc.). For comparability, at all steps we included participants with complete data only (N=5023). Figure 6 presents the outcome of this analysis.

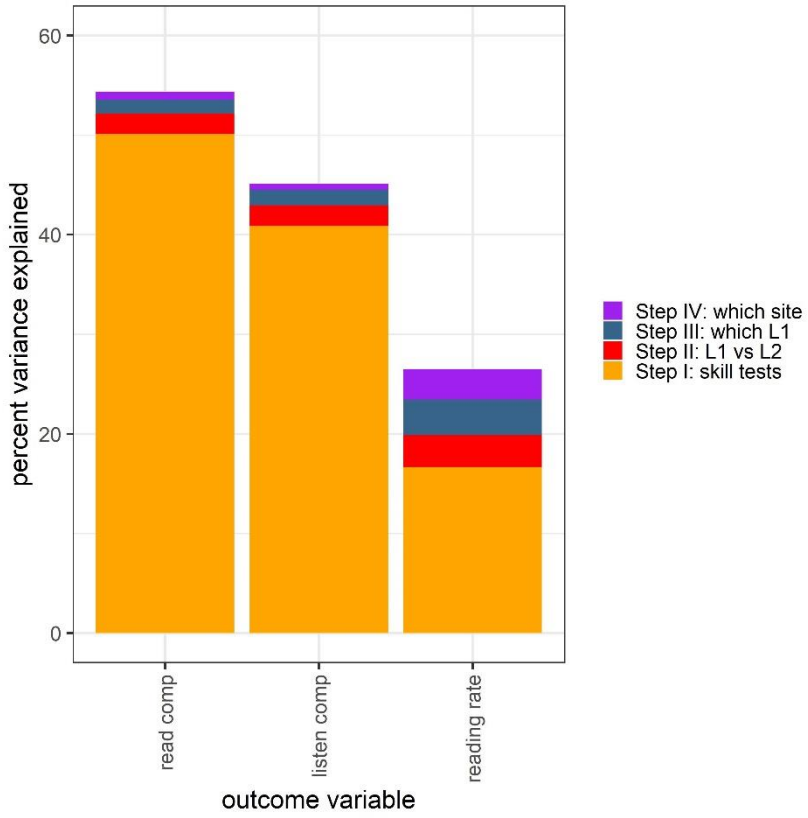

Figure 6. Stepwise partitioning of variance in English reading comprehension (read comp), listening comprehension (listen comp) and reading rate. Step 1 - component skills of English; Step 2 - differences between native and non-native speakers; Step 3 - differences between L1s (other than English); Step 4 - inter-site variability within a country and native language.

In analyses of reading and listening comprehension accuracy (two left columns in Figure 6), the vast majority of variance was explained at Step 1, by the English component skills of listening and reading (the latter also included listening comprehension as a component skill). Thus, the absolute amount of variance explained by component skills was 50.1% in reading comprehension and 40.9% in listening comprehension. These estimates amount to 91-92% of the total variance explained by all variables at Steps 1-4. The inclusion of the L1-L2 status variable at Step 2 added little explained variance in reading and listening comprehension accuracy (ΔR^2=2.1% for both reading and listening comprehension, which amounts to 3.8% and 4.5% of

total variance at Steps 1-4 for reading and listening comprehension, respectively). Similarly, at Steps 3 and 4, we found that sample characteristics (i.e., participants' L1 and the site within a country and L1) led to little improvements in variance explained: all ΔR^2s$\leq$1.6%, which constitute a relative contribution of 3.5% or less of the total variance explained[4]. Notably, the added variance explained in Steps 3 and 4 was similar to amount of variance associated with L1-L2 differences (Step 2).

A different picture emerged when reading rate served as the dependent variable (rightmost column of Figure 6). First, the total variance explained by variables at Steps 1-4 was lower (R^2=26.5% in total). Most likely, this reduction in R^2 is related to the lesser impact of English component skills (R^2 = 16.7% at Step 1 for reading rate, less than half than in reading and listening comprehension). Yet, in addition to the absolute decrease in R^2, component skills also played a lesser *relative* role in predicting reading rate, accounting for 62.9% of the total variance explained at Steps 1-4. Instead, larger relative portions of variance were explained by the L1-L2 distinction (ΔR^2=3.2%, accounting for 12.0% of total variance explained) and by L1 background (ΔR^2=3.6%, accounting for 13.5% of total variance explained). Cross-site differences within countries and native languages (Step 4) accounted for an additional variance at a level comparable to that explained by Steps 2 and 3: ΔR^2=3.1%, i.e., 11.6% of total variance. We discuss the findings of the partitioning approach in the General Discussion, below.

We note that the order in which variables were entered into analyses above was meant to provide information regarding the impact of the L1-L2 distinction and other participant-characteristics *beyond* the impact of individual differences in component skills (i.e., after this variance is controlled for in Step 1). However, alternative orders can be used to examine related questions. One such alternative ordering of variables, putting the L1-L2 distinction before

component skills, is explored in Appendix S8. Crucially, this analysis still replicated the key finding above regarding the minor impact of the L1-L2 distinction (vs. component skills) on reading and listening comprehension. Also mirroring the results above, the L1-L2 distinction had a stronger link to reading rate, one that under the alternative order was close in size to the portion of variance explained by component skills. See Appendix S8 for details.

4. General Discussion

Research into L2 reading is in strong need for large data sources that afford a broad coverage of language backgrounds and component skills of reading, and also provide cross-sample consistency and comparability in design, administration, apparatus and samples of participants (see bibliometric analysis in Kuperman et al., 2022, and the Introduction). The first goal of this paper was to introduce to the research community a new data source – the ENglish Reading Online (ENRO) – that fits this description. The ENRO database contains data from 7338 university students, representing 30 recruitment sites, 19 countries and 16 unique dominant or primary languages of instruction. This coverage is on par with the most inclusive meta-analyses currently available in the field. Many of the participants reported English as their first language or as a language they acquired before the age of 5: Under this criterion, discussed in detail below, they were considered L1 speakers of English (see Methods for the imputation procedure). Most other participants were advanced learners of English who passed English language examinations to be accepted to an educational institution.

The data include participant-level performance in a text reading task (reporting measures of comprehension accuracy and reading rate), a listening for comprehension task, as well as seven tests of component skills selected to represent major contributors to English reading

proficiency identified in the literature (e.g., vocabulary, spelling, orthographic and grammar knowledge, and lexical decision). Trial-level data from all tasks are made publicly available. Furthermore, detailed questionnaires provide rich data on language background and use as well as demographic characteristics of participants (Gullifer & Titone, 2020). As envisioned by the study design, the ENRO data make possible both the "big picture" exploratory studies of L1 and L2 proficiency and targeted investigations of language and reading behavior driven by specific theoretical questions or done on subsets of language backgrounds, participants or items. This paper confined itself to description and methodological validation of the data collected and – as its second goal – focused on only a few "big picture" questions. The questions, broached in detail in the Introduction, were (1) how similar or different is language and reading behavior of L1 and L2 readers of English and (2) what predictors explain variance in measurable outcomes of English reading and listening comprehension. Below we summarize the outcomes of the validation and major findings.

4.1 Methodological foundations: reliability and validity of ENRO measures

ENRO assessments in the full sample of participants showed high reliability in all outcome measures, and reliability estimates were especially high for reading and listening tasks. Similarly high reliability was observed in L1 and L2 readers of English when considered separately. This suggests high stability and utility of the present data for investigation of individual differences. Furthermore, the validity of the ENRO data was supported by the correlations between ENRO measures and its compatibility with the results of recent meta-analyses of reading and listening comprehension (Jeon & Yamashita, 2014; 2022; In'nami et al., 2022). In particular, the rank order of component skills of reading and listening comprehension,

ordered by the strength of the correlation between the skill test and the comprehension outcome, was highly comparable between the meta-analyses and the present primary study (see Appendix S7). Together, these analyses show that the present data source both provides high-quality data and can obviate the shortcomings present in current studies (which also limits extant meta-analyses): e.g., heterogeneity of studies and samples, lack of information on the correlations between the different predictors, and lack of direct comparison of L2 with L1 participants. The remainder of the General Discussion elaborates on theoretical goals of the paper.

4.2 How does language and reading behavior compare across L1 and L2 English speakers?

This question is central for the present paper and stems from the bias that existing studies in the field have towards emphasizing differences rather than similarities between L1 and L2 speakers of a language under examination (see Introduction). As we argue in the Introduction, the bias is at least partially grounded in the null hypothesis-testing logic, which is designed to detect how much the groups differ from each other but not how much they overlap. As a result, statistically significant but practically unimportant differences associated with the L1-L2 distinction occupy a disproportionally large place in the literature. This paper offers several quantitative tests of the degree of overlap vs. difference between L1 and L2 populations. The goal is to provide empirical grounding to the question of how fruitful it is to adhere with the L1 vs. L2 binary distinction rather than consider English language and reading proficiency across the continuum of skill, ability, and experiences which spans over L1 and L2 speakers alike (see Diependaele et al., 2013; Gullifer et al., 2021).

The present data show overwhelming evidence for similarities, rather than differences, between L1 and L2 in their reading behavior and relative contributions of component skills of

English reading to this behavior. Perhaps the most telling finding in this regard is the very small difference in reading and listening comprehension levels between L1 and L2 speakers (Cohen's *d* = .00 and .22, respectively; Table 3). Not all component skills of English comprehension showed similarly small differences in L1 vs. L2 performance: The differences were substantial in, for example, tests of grammar knowledge (*d* =1.12), lexical decision accuracy in LexTALE (*d*=1.01), and vocabulary knowledge (*d* = .92; see Table 3). Yet our correlational analyses determined that – despite group differences in mean performance – the skills measured by these tests played a similarly strong role in predicting reading and listening comprehension and reading rate both in L1 and L2 samples. Taken together, these findings indicate that L1 and (advanced) L2 speakers of English attain similar levels of comprehension, and that the relative roles played by multiple component skills in this attainment are similar as well.

The latter observation finds further support in correlational and factor analyses (section 3.4), which examined interrelations between predictors and outcomes of comprehension tasks. Magnitudes of the correlations between ENRO test scores were highly similar between L1 and L2 participants, so much so that the vast majority of differences in the correlation strength that reached statistical significance were too small to be practically important. Not only did individual component skills of reading exercise a similar influence on English reading and listening comprehension and reading rate in L1 and L2 groups, but also relations between those individual component skills were highly comparable across the groups. The exploratory factor analysis went beyond pairwise correlations to determine how the tests we administered grouped together to represent common latent variables. A comparison of resulting factor solutions for L1 and L2 participants revealed, again, highly overlapping results. Both solutions indicated three factors representing untimed component skills of English proficiency (e.g., vocabulary,

grammar, orthography, and spelling knowledge, as well as accuracy in lexical decision tasks); timed responses (lexical decision RTs); and comprehension scores (reading and listening). In both exploratory factor solutions, one of the factors that reading rate loaded on was comprehension. Faster readers showed lower comprehension levels, i.e., a clear-cut case of the speed-accuracy trade-off (Heitz, 2014; Mulder et al., 2021). Some subtle differences between L1 and L2 factor solutions emerged as well: e.g., only in the L2 sample listening and reading comprehension and reading rate loaded on the first factor ("untimed responses"). Indeed, this difference may point to some subtle differences between the two samples in terms of the latent structure of English proficiency profiles, e.g., that in L2 readers of English, comprehension and reading rate are more closely related to component skills than in L1 readers.[5] Still, we contend that the differences were minor compared to the overlap in the inter-relations between component skills and outcomes of English reading comprehension and their attribution to latent constructs.

Lastly, the partitioning-of-variance analysis (section 3.5), showed that the contribution of the L1-L2 contrast to explaining variance in main outcomes of English comprehension tasks (accuracy of listening and reading comprehension, reading rate) was minor, accounting for a 2-3% increase in the amount of explained variance in all cases. In fact, this magnitude of contribution was on par with the contributions to explained variance associated with the site within the country where data collection took place. In other words, English performance differences between students attending different universities within a country are comparable with the differences associated with the L1-L2 contrast. In contrast, the vast majority of explained variance in reading and listening comprehension (over 90%) and reading rate (63%) is traced back to the individual performance in component skills of English proficiency. This

finding further puts into perspective how limited the practical impact of the L1-L2 contrast is, despite its salient role as a theoretical construct. Categorical distinctions (either between specific L1s reported by L2 speakers of English in our dataset, or the binary L1- L2 distinction) are overshadowed by the individual mastery of component skills of reading in English as predictors of reading and listening comprehension in this language (e.g., Kuperman et al., 2022; Nisbet et al., 2022). This suggests that a fruitful approach for further studies of university-level advanced learners of English concentrates on the shared nature of language and reading acquisition and knowledge rather than the demonstrably small differences.

4.3 What explains variance in reading and listening comprehension and reading rate?

As noted above, in contrast to the minor impact of the L1-L2 distinction, component skills explained most of the variance in both English comprehension accuracy and rate measures, leaving little explanatory power not only to that binary distinction, but also to differences between specific L1 backgrounds (for L2 speakers of English) and within-country differences. Correlational analyses allowed a further insight into which specific skills were predictive of these outcomes. They highlighted the strong role of the same higher-order skill set as indicated in meta-analyses of L2 comprehension (Jeon & Yamashita, 2014; 2022; In'nami et al., 2022): i.e., grammar and vocabulary knowledge (also measured in lexical decision tasks in our data) and listening comprehension. Thus, our data confirm and enrich the current understanding of how component skills are coordinated and relied on to achieve reading and listening comprehension: This way is demonstrably highly similar in L1 and advanced L2 speakers of English.

Furthermore, we observed consistent differences between predictors of reading and listening comprehension on the one hand, and reading rate, on the other. These differences –

emphasized earlier by Busby and Dahl (2021), Dirix et al., (2020) and Kuperman et al. (2022), among others – emerge in all analyses reported above. First, descriptive statistics and visualizations of test performance (see Figures 1 and 2, and Appendix S5) revealed a strong dispersion of mean reading rates, with a clear distinction in performance between L1 and L2 speakers of English (Cohen's *d*=.68). This distinction was not observed in either reading or listening comprehension data. In correlational analyses, reading rate was predicted most strongly by other chronometric measurements, including RTs in lexical decision tasks and in the timed segmentation task. This contrasts with the hierarchy of predictors for comprehension accuracy, outlined above. Exploratory factor analysis revealed a degree of convergence between reading rate and reading and listening comprehension, as they loaded on the same factor in both L1 and L2 factor solutions. Finally, the partitioning-of-variance analysis showed a much smaller total amount of variance explained in reading rate (26.5%) compared to reading (50.1%) and listening (40.9%) comprehension tasks. The relative contribution of component skills was smaller too, and the L1-L2 contrast explained relatively more variance in reading rate as opposed to comprehension tasks, see Kuperman et al. (2022) for similar findings.

Considered jointly, these findings indicate substantial dissociation between reading comprehension and reading rate as hypothesized facets of reading proficiency, even if the populations we consider are statistically matched in their reading and listening comprehension performance (see also Vander Beken, 2020). It may be worthwhile for future research to ask if reading comprehension and reading speed (i.e., the quality of knowledge and fluency) should be treated as distinct dimensions of reading, being influenced by different developmental factors and relying on largely different skills and abilities. As Kuperman et al. (2022) argue based on within-participant comparisons of eye-movements and reading rates in L1 and L2, reading rate –

unlike reading comprehension – may be sensitive to domain-general skills, including cognitive speed. For educational research, these findings are noteworthy, since they indicate that achieving a native-like performance in the quality of comprehension among advanced learners of English does not come with the native-like reading speed. Yet speed is of obvious importance for workplace and academic environments, which often posit strict time limits for tasks, including examinations (e.g., Dirix et al., 2020). Thus, an additional focus on reading speed may be a worthwhile priority for instructional programs for L2 learning. For researchers, these findings highlight the importance of shifting the attention from the current focus on reading and listening comprehension towards the much less studied topic of fluency of reading. Our data show that fluency (measured as reading rate) is a source of much greater variability than comprehension even in advanced L2 learners, while causes of that variability and even its direction – slower is better – are not yet entirely understood.

4.4 Limitations and future directions

The present body of findings needs to be interpreted while keeping in mind the nature of our populations, tasks and operational definitions. We note that our L2 speakers are mostly advanced university-level speakers of English, often with early and intensive exposure to English. Also, for simplicity we gloss over the distinction between L2 speakers studying in English-dominant vs. non-English-dominant institutions (e.g., ESL vs. EFL), leaving it to future research (see de Cat et al., 2022; Tiv et al., 2022). Furthermore, two operationalizations that we adopted are perhaps most relevant to the interpretation of results. First, we defined L2 speakers of English as those who acquired English at or after the age of 5, which is a common age of formal schooling in many participating countries. While adopted by some researchers as an operationalization of the

L1/L2 distinction, this threshold is not universal: e.g., some research groups define as L2 speakers those individuals who started acquiring English after the first year of life. Particularly relevant for samples collected in English-dominant countries, selection of the threshold age may affect the strength of a contrast between groups defined as L1 vs L2 speakers of English – to the degree that a person exposed to English since, say, the age of 1 differs from that exposed to English since the age of 5. Our second, related design decision was to impute the L1 and L2 status for some participants based on the dominant or primary language of instruction in the respective institution: All participants in L2 sites were labeled as L2 speakers of English (regardless of self-reports), and in L1 sites where the L1-L2 distinction in the sample was too small (less than 5%) self-reporting L2 participants were relabeled as L1s. Availability of ENRO data and code, including in particular the rich language background data collected, enable researchers to validate the present findings against alternative and more fine-grained definitions of L1 or L2 speakers of English.

Further limitations are related to design choices we had to make. The web-based nature of the study and time constraints led us to exclude from the battery some important component skills of English (e.g., phonological and morphological awareness) and general cognitive measures known to correlate with L2 proficiency (e.g., working memory). ENRO also lacks the inclusion of L1 tests (for L2 speakers of English), which would allow to compare L2 reading rate and comprehension to L1 reading rate and comprehension. The reason why we did not include this aspect is that it is highly taxing to ensure equivalent tests given the large number of L1s in the ENRO sample. We therefore chose to leave this comparison for future studies, which can focus, e.g., on within-L1 analyses of L1-L2 reading comprehension and rate in specific languages of interest.

The findings of the present analyses are further limited to a specific text genre and a specific type of comprehension questions. Expository (encyclopedia style) texts are more likely to benefit from slower, careful reading than, say, fiction. Similarly, the use of multiple choice questions as a measure of comprehension may have masked differences between L1 and L2 readers in the richness and degree of organization of text memory. It is known that recall questions are more difficult to answer than recognition questions, and a number of studies have suggested that there may be more differences between L1 and L2 readers in recall than in recognition (Li & Kirby, 2015; Vander Beken et al., 2020).

A last limitation is that we only examined English as a target language, a language that is already massively over-represented in second language research (see the Introduction). We advocate for the creation of similar data resources with target languages other than English.

Despite the limitations, the ENRO project provides a rich database that enables multiple lines of investigation, far exceeding the first analytical pass on the data presented so far. In this section, we review some directions that we consider to be of theoretical interest for future research using the ENRO data. First, we note that ENRO data come with a rich questionnaire tapping into ecology of language use (e.g., the frequency and nature of using each spoken and read language in various settings) and self-reported measures of proficiency and age-of-acquisition of English speech and reading, along with many other demographic characteristics (Gullifer & Titone, 2020). Following prior work on the impact of language background and use on individual performance (Gullifer & Titone, 2021a; Pivneva et al., 2014; Titone et al., 2011; Tiv et al., 2022; Vingron et al., 2021), we encourage researchers to make use of this rich data.

Second, the present analysis focused on mean individual performance in each task. Yet we also make available trial-level data for all tasks (where applicable). This reporting makes

possible an in-depth investigation of, say, lexical decision latencies and accuracy as a function of the participant's proficiency in language, their demographic characteristics, and various word-level properties (Gullifer & Titone, 2021b). Another potential avenue using the trial-level data is analysis of measurement invariance, estimating the extent to which ENRO tests measure the same constructs in L1 and L2 speakers (Luong & Flake, 2022).

Third, the ENRO data give access to L1 backgrounds that vary widely in their writing systems and linguistic properties of the oral language. This paves the way for a systematic study of the influence that the linguistic distance and the script distance between L1 and English as L2 has on L2 learning and proficiency (e.g., Schepens et al., 2013; Wichman et al., 2010). We note that in the current population this influence is likely to be minor given the small group differences between L1 and L2 speakers observed.

Fourth, two samples in our data were collected via crowdsourcing platforms. Comparing their results against university-based samples representing similar languages would be of methodological interest for the quality of data in online-administered tasks that crowdsourcing can provide relative to student samples recruited from university convenience pools.

5. Concluding remarks

This paper introduces the ENRO project as a high-power source of data on English reading and its component skills obtained from over 7,000 speakers of English from diverse L1 backgrounds. The project further presents rich meta-data on demographic characteristics of participants as well as detailed contextual information about their use of spoken and written languages. Uniform parameters of data collection and selection of participants (university students, either L1 or advanced L2 speakers of English) and demonstrable high reliability of the tests contribute to the

utility of the data for both the bird's-eye-view comparison of large groups of participants and the study of individual differences. We conducted analyses that addressed questions often posited as central in the L2 research and outlined some of the many future directions that can be pursued by further mining the ENRO data. It is our hope that the large-scale empirical base provided by the ENRO project and similar mega-studies will help to expand the scope, depth, and methodological consistency of the inquiry into reading behavior.

Endnotes

[1] Two noteworthy cases are McGill University, where English was assigned as a source language because it is an English-language institution in the predominantly French-speaking Canadian province of Quebec, and Université Libre de Bruxelles, where French was assigned as a source language given that it is a French-language institution in Belgium. We retained Arabic as the source language of Arabic-speaking students in the Hebrew University of Jerusalem (where Hebrew is the language of instruction), to distinguish between samples of native Hebrew and Arabic speakers in that university.

[2] Because it is based on one trial only, no reliability estimates were obtained for the text segmentation task.

[3] Consistent with our definition of L1-L2 status above, a native speaker of a given language was defined as someone who learned the language before the age of 5. Participants could have multiple languages coded as their native languages. For simplicity, in this analysis we only included languages that met our criterion of a native language in 10 participants or more. The final list included 34 dummy-coded variables, according to the languages chosen/entered by participants (note that some of these were entered by participants as free text): Albanian, Arabic, Bengali, Cantonese, Chinese, Creole, Croatian, Dutch, Farsi, French, German, Gujarati, Hebrew, Hindi, Italian, Japanese, Korean, Malayalam, Mandarin, Polish, Portuguese, Punjabi, Romanian, Russian, Serbian, Spanish, Tagalog, Tamil, Telugo, Thai, Turkish, Ukrainian, Urdu, and Vietnamese.

[4] Despite the small amount of associated variance, all increases in explained variance at Steps 2-4 constitute a significant improvement in model fit (all *p*s < .05). This is expected given ENRO's large sample size.

[5] The Factor Analysis in Table 6 above reveals additional subtle differences between the samples: For example, Table 6 shows a positive loading of reading rate on Factor 1 only in the L2 sample. However, such apparent differences between samples have to do with the cut-off used to flag strong loadings (> |0.3|): In this particular case, for instance, the same positive loading of reading rate onto Factor 1 exists in L1s, only it is estimated at a sub-threshold value of 0.23 (compared to 0.35). See the code on OSF for full factor analysis output. Confirmatory factor analysis and structural equation modelling will be better techniques to answer these questions.

References

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual review of psychology*, *63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

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*, 1-10. https://doi.org/10.1162/opmi_a_00054

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*. Available online 14 October 2022. https://doi.org/10.1016/j.tics.2022.09.015

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of memory and language, 109, 104047.* https://doi.org/10.1016/j.jml.2019.104047

Brysbaert, M., & Rastle, K. (2021). *Conceptual and Historical Issues in Psychology*. Pearson Higher Ed.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441. https://psycnet.apa.org/doi/10.1037/xhp0000159

Busby, N. L., & Dahl, A. (2021). Reading Rate of Academic English Texts: Comparing L1 and Advanced L2 Users in Different Language Environments. *Nordic Journal of English Studies*, *20*(1). http://doi.org/10.35360/njes.542

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods, 49*(2), 602-615. https://doi.org/10.3758/s13428-016-0734-0

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Science, 9(3)*, 33. https://doi.org/10.3390/bs9030033

De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2022). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism: Language and Cognition,* 1-13. https://doi.org/10.1017/S1366728922000359

DeKeyser, R. M. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition*. *An introduction* (3rd ed., pp. 83-104). New York, NY: Routledge. http://dx.doi.org/10.4324/9780429503986-5

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*(5), 843-863. https://doi.org/10.1080/17470218.2012.720994

Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly, 55*(3), 371-397. https://doi.org/10.1002/rrq.277

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading research quarterly, 38*(1), 78-103. https://doi.org/10.1598/RRQ.38.1.4

Foorman, B. R., Petscher, Y., & Herrera, S. (2018). Unique and common effects of decoding and language factors in predicting reading comprehension in grades 1–10. *Learning and Individual Differences, 63,* 12-23. https://doi.org/10.1016/j.lindif.2018.02.011

Godfroid, A. (2020). *Eye tracking in Second Language Acquisition and bilingualism. A research synthesis and methodological guide*. New York: Routledge. https://doi.org/10.4324/9781315775616

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6-10. https://doi.org/10.1177/074193258600700104

Grabe, W., & Stoller, F. L. (2019). Teaching and Researching Reading: Third Edition. Routledge. https://doi.org/10.4324/9781315726274

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*(1), 163-189. https://doi.org/10.1146/annurev.psych.48.1.163

Grosjean, F. (2008). Studying bilinguals. Oxford University Press.

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition, 23(2),* 283-294. https://doi.org/10.1017/S1366728919000026

Gullifer, J., & Titone, D. (2021a). Bilingualism: A Neurocognitive Exercise in Managing Uncertainty. *Neurobiology of Language*, *2(4)*: 464–486. https://doi.org/10.1162/nol_a_00044

Gullifer, J. W., & Titone, D. (2021b). Engaging proactive control: Influences of diverse language experiences using insights from machine learning. *Journal of Experimental Psychology: General, 150*(3), 414–430. https://psycnet.apa.org/doi/10.1037/xge0000933

Gullifer, J., Kousaie, S., Gilbert, A., Grant, A., Giroud, N., Coulter, K., . . . Titone, D. (2021). Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. *Applied Psycholinguistics, 42*(2), 245-278. https://doi.org/10.1017/S0142716420000521

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, *8*, 150. https://doi.org/10.3389/fnins.2014.00150

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160. https://doi.org/10.1007/BF00401799

In'nami, Y., Koizumi, R., Jeon, E. H., Arai, Y. (2022). L2 speaking and its external correlates. In: Jeon, E.H. & In'nami, Y. *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. John Benjamins. pp. 235—284. https://doi.org/10.1075/bpa.13.11jeo

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language learning, 64*(1), 160-212. https://doi.org/10.1111/lang.12034

Jeon, E. H., & Yamashita, J. (2022). L2 reading comprehension and its correlates: An updated meta-analysis. In: Jeon, E.H. & In'nami, Y. *Understanding L2 Proficiency: Theoretical and meta-analytic investigations*. John Benjamins. pp. 29—86. https://doi.org/10.1075/bpa.13.03jeo

Just, M.A., Carpenter, P. A. (1987). *Speedreading: The Psychology of Reading and Language Comprehension*. Newton, MA: Allyn & Bacon.

Kaushanskaya, M., Blumenfeld, H., & Marian, V. (2020). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition, 23(5)*, 945-950. https://doi.org/10.1017/S1366728919000038

Kim, Y. S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading, 21*(4), 310-333. https://doi.org/10.1080/10888438.2017.1291643

Kuperman, V., Kyröläinen, A. J., Porretta, V., Brysbaert, M., & Yang, S. (2021). A lingering question addressed: Reading rate and most efficient listening rate are highly similar. *Journal of Experimental Psychology: Human Perception and Performance, 47(8),* 1103—1112. https://psycnet.apa.org/doi/10.1037/xhp0000932

Kuperman, V., Siegelman, N., Schroeder, S., Acartürk, C., Alexeeva, S., Amenta, S., ... & Usal, K. A. (2022). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, 1-35. https://doi.org/10.1017/S0272263121000954

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics, 36(5),* 611-634. https://doi.org/10.1093/applin/amu007

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition, 23(5)*, 938-944. https://doi.org/10.1017/S1366728918001153

Luong, R., & Flake, J. K. (2022). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and

reporting. *Psychological Methods*. Advance online publication. https://psycnet.apa.org/doi/10.1037/met0000441

McCarron, S. P., & Kuperman, V. (2022). Effects of year of post-secondary study on reading skills for L1 and L2 speakers of English. *Journal of Research in Reading, 45*(1), 43-64. https://doi.org/10.1111/1467-9817.12380

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—433. https://doi.org/10.1037/a0033890

Mulder, E., van de Ven, M., Segers, E., Krepel, A., de Bree, E. H., de Jong, P. F., & Verhoeven, L. (2021). Word-to-text integration in English as a second language reading comprehension. *Reading & Writing, 34(4)*, 1049–1087. https://doi.org/10.1007/s11145-020-10097-3

Nisbet, K., Bertram, R., Erlinghagen, C., Pieczykolan, A., & Kuperman, V. (2022). Quantifying the difference in reading fluency between L1 and L2 readers of English. *Studies in Second Language Acquisition*, *44*(2), 407-434. https://doi.org/10.1017/S0272263121000279

Peng, P., Fuchs, D., Fuchs, L. S., Elleman, A. M., Kearns, D. M., Gilbert, J. K., ... & Patton III, S. (2019). A longitudinal analysis of the trajectories and predictors of word reading and reading comprehension development among at-risk readers. *Journal of Learning Disabilities*, *52*(3), 195-208. https://doi.org/10.1177/0022219418809080

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook*. Blackwell Publishing, pp. 227–247. https://psycnet.apa.org/doi/10.1002/9780470757642.ch13

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22-37. https://doi.org/10.1080/10888438.2013.827687

Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 787—796. https://doi.org/10.1037/a0035583

Raudszus, H., Segers, E., & Verhoeven, L. (2021). Patterns and predictors of reading comprehension growth in first and second language readers. *Journal of Research in Reading*, *44*(2), 400-417. https://doi.org/10.1111/1467-9817.12347

Revelle, W. (2015). Package 'psych'. *The comprehensive R archive network*, *337*, 338.

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, *16*(4), 744-755. https://psycnet.apa.org/doi/10.1177/1745691620966795

Schepens, J., Van der Slik, F., Van Hout, R., Borin, L., & Saxena, A. (2013). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In L. Borin, A. Saxena (Eds.), *Approaches to measuring linguistic differences*, De Gruyter Mouton, Berlin (2013), pp. 199-230. https://doi.org/10.1515/9783110305258

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. *Journal of Research in Personality*, *47*(5), 609-612. https://doi.org/10.1016/j.jrp.2013.05.009

Segalowitz, N. (2010). Cognitive bases of second language fluency. London, England: Routledge. https://doi.org/10.4324/9780203851357

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*, 1-21. https://doi.org/10.3758/s13428-021-01772-6

Sommers, M. S., Hale, S., Myerson, J., Rose, N., Tye-Murray, N., & Spehar, B. (2011). Listening comprehension across the adult lifespan. *Ear and Hearing*, *32*(6), 775-781. 10.1097/AUD.0b013e3182234cf6

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271.

Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism: Language and Cognition*, *22*(2), 401-415. https://doi.org/10.1017/S1366728917000682

Tiv, M., Kutlu, E., Gullifer, J. W., Feng, R. Y., Doucerain, M. M., & Titone, D. A. (2022). Bridging interpersonal and ecological dynamics of cognition through a systems framework of bilingualism. *Journal of Experimental Psychology: General , 151*(9), 2128–2143. https://doi.org/10.1037/xge0001174

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–

L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1412—1431. https://psycnet.apa.org/doi/10.1037/a0024492

Tye-Murray, N., Sommers, M., Spehar, B., Myerson, J., Hale, S., & Rose, N. S. (2008). Auditory-visual discourse comprehension by older and young adults in favorable and unfavorable conditions. *International Journal of Audiology*, *47*(sup2), S31-S37. https://doi.org/10.1080/14992020802301662

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458-483. https://doi.org/10.1006/jmla.1998.2584

Verhoeven, L., & Perfetti, C. (Eds.). (2017). *Learning to read across languages and writing systems*. Cambridge University Press. https://doi.org/10.1017/9781316155752

Vingron, N., Palma, P., Gullifer, J. W., Whitford, V., Friesen, D., Jared, D., & Titone, D. (2021). What Are the Modulators of Cross-Language Syntactic Activation During Natural Reading?. *Frontiers in Communication*, 112. https://doi.org/10.3389/fcomm.2021.597701

Vander Beken, H., De Bruyne, E., & Brysbaert, M. (2020). Studying texts in a non-native language: a further investigation of factors involved in the L2 recall cost. *Quarterly Journal of Experimental Psychology*, *73*(6), 891-907. https://doi.org/10.1177/1747021820910694

Vermeiren, H., Vandendaele, A., & Brysbaert, M. (2022). Validated tests for language research with university students whose native language is English: Tests of vocabulary, general knowledge, author recognition, and reading comprehension. *Behavior Research Methods*, 1-33. https://doi.org/10.3758/s13428-022-01856-x

Whitford, V., & Joanisse, M. F. (2021). Eye Movement Measures of Within-Language and Cross-Language Activation During Reading in Monolingual and Bilingual Children and Adults: A Focus on Neighborhood Density Effects. *Frontiers in psychology, 12,* 674007. https://doi.org/10.3389/fpsyg.2021.674007

Whitford, V., & Titone, D. (2019). Lexical entrenchment and cross-language activation: Two sides of the same coin for bilingual reading across the adult lifespan. *Bilingualism: Language and Cognition, 22(1),* 58-77. https://doi.org/10.1017/S1366728917000554

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Number of participants with valid data in each measure.

Appendix S2: Additional information regarding sample units.

Appendix S3: Description of tests of component skills.

Appendix S4: Details regarding texts in reading listening comprehension tasks.

Appendix S5: Descriptive plots for tests of component skills.

Appendix S6: Additional correlation tables.

Appendix S7: Comparison of correlations between ENRO and meta-analytic estimates.

Appendix S8: An alternative order of variables in partitioning of variance analysis.