LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Rethinking First Language–Second Language Similarities and Differences in English Proficiency: Insights From the ENglish Reading Online (ENRO) Project

Noam Siegelman [a,b] Irina Elgort [c] Marc Brysbaert,[d] Niket Agrawal,[e] Simona Amenta,[f] Jasmina Arsenijević Mijalković,[g] Christine S. Chang,[h] Daria Chernova,[i] Fabienne Chetail,[j] A. J. Benjamin Clarke,[k] Alain Content [j] Davide Crepaldi,[l] Nastag Davaabold,[m] Shurentsetseg Delgersuren,[m] Avital Deutsch,[a] Veronika Dibrova,[n] Denis Drieghe,[o] Dušica Filipović Đurđević,[g,p] Brittany Finch [q] Ram Frost,[a] Carolina A. Gattei,[r] Esther Geva,[s] Aline Godfroid,[q] Lindsay Griener,[t] Esteban Hernández-Rivera,[u] Anastasia Ivanenko,[v] Juhani Järvikivi,[t] Lea Kawaletz,[w] Anurag Khare,[e] Jun Ren Lee,[h] Charlotte E. Lee,[o] Christina Manouilidou,[x] Marco Marelli,[f] Timur Mashanlo,[n] Ksenija Mišić,[g] Koji Miwa,[y] Pauline Palma,[u] Ingo Plag,[w] Zoya Rezanova,[n] Enkhzaya Riimed,[m] Jay Rueckl,[b,z] Sascha Schroeder,[aa] Irina A. Sekerina [bb] Diego E. Shalom,[cc,r] Natalia Slioussar [v,i] Neža Marija Slosar,[x] Vanessa Taler,[dd,ee] Kim Thériault,[dd] Debra Titone,[u] Odonchimeg Tumee,[m] Ross van de Wetering,[c] Ark Verma,[e] Anna Fiona Weiss [ff] Denise Hsien Wu,[h] and Victor Kuperman[gg]

[a]The Hebrew University of Jerusalem [b]Haskins Laboratories [c]Victoria University of Wellington [d]Ghent University [e]Indian Institute of Technology – Kanpur [f]University of Milano-Bicocca [g]University of Belgrade [h]National Taiwan Normal University [i]Saint Petersburg State University [j]Université libre de Bruxelles [k]Thammasat University [l]International School for Advanced Studies [m]Khovd Branch of National University of Mongolia [n]Tomsk State University [o]University of Southampton [p]University of Novi Sad [q]Michigan State University [r]Universidad

DOI: 10.1111/lang.12586

Torcuato Di Tella [s]University of Toronto [t]University of Alberta [u]McGill University [v]Higher School of Economics Moscow [w]Heinrich-Heine-Universität Düsseldorf [x]University of Ljubljana [y]Nagoya University [z]University of Connecticut [aa]University of Goettingen [bb]City University of New York [cc]Universidad de Buenos Aires [dd]University of Ottawa [ee]Bruyère Research Institute [ff]Catholic University of Eichstaett-Ingolstadt [gg]McMaster University

Abstract: This article presents the ENglish Reading Online (ENRO) project that offers data on English reading and listening comprehension from 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 first (L1) and

A one-page Accessible Summary of this article in non-technical language is freely available in the online Supporting Information and at https://oasis-database.org

All data are available through the Open Science Framework repository (https://osf.io/gzyqf). Other than the first three and the last author, the order of authors is alphabetical. 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. Research reported in this publication was supported by the following grants: a Social Sciences and Humanities Research Council of Canada Partnered Research Training Grant, 895-2016-1008 (PI: G. Libben); a Canada Research Chair (Tier 2; PI: V. Kuperman); a Social Sciences and Humanities Research Council of Canada Insight Grant, 435-2021-0657 (PI: V. Kuperman); an Azrieli Early Career Faculty Fellowship (PI: N. Siegelman); a Natural Sciences and Engineering Research Council of Canada Discovery Grant (PI: D. Titone); the Russian Science Foundation, #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; the Israel Science Foundation, #705/20 (PI: R. Frost); a Faculty Research Grant, Faculty of Humanities and Social Sciences, Victoria University of Wellington, #226239 (PI: I. Elgort); a Professional Staff Congress-City University of New York Award, #64464-00-52 (PI: I. A. Sekerina); the Chinese Language and Technology Center of National Taiwan Normal University (PI: Y. T. Sung); the Tomsk State University Development Programme (Priority2030); the Italian Ministry of Education and Research, #2017W8HFRX (PIs: V. Pirrelli & D. Crepaldi); and the Eunice Kennedy Shriver National Institute of Child Health and Human Development, #HD091013 (PI: D. Compton).

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

The handling editors for this manuscript were Judit Kormos and Pavel Trofimovich.

second (L2) language 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 for how researchers can mine ENRO data for future studies.

Keywords reading; second language proficiency; bilingualism; cross-linguistic research; open science

Introduction

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

Although research on acquisition and use of a second language (L2) is growing both in the scope of topics and in the number of publications (see bibliometric analysis in Kuperman et al.'s study, 2023), research is still fairly limited in its coverage of languages studied both as participants' first languages (L1s) and as 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"). When this coverage is put in perspective, it accounted for roughly one-third of the diversity of L1 backgrounds among undergraduate students at the midsized Canadian McMaster University (enrolment of 30,000) and about one-half of the languages taught at that university as L2s. Limited coverage aside, existing studies have often used different assessments, measurement procedures, and inclusion/exclusion criteria (de Cat et al., 2022; Surrain & Luk, 2019). This variability has reduced the comparability of findings across studies (de Bruin, 2019).

Another bias in existing studies has been that they have tended to focus on differences between L1 and L2 processing rather than on L1 and L2 similarities, even if few researchers would endorse 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, the null hypothesis, has been the engine of scientific progress. Null-hypothesis testing is by definition geared toward finding differences between groups and conditions rather than similarities and requires a minimum of 860 participants to test the null effect (Cohen's $d < 0.2$) between two groups and four times this number to test the absence of an interaction ($d = 0.2$ vs. $d = 0.0$) between a between-groups variable and a repeated measures variable (Brysbaert, 2019). Since such a sample size is hard to achieve, practicalities of research have dictated a focus on group differences rather than on similarities between L1 and L2 reading. Thus, a chapter 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 versus one page outlining similarities. A similar (implicit) bias has emerged in meta-analyses of L2 reading or listening comprehension. By omitting L1 speakers from consideration, such meta-analyses can only speak to the possible differences between L1 and L2 speakers rather than to their potential similarities.

Importantly, a bias toward 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 overemphasize L1–L2 differences can be found in research on cross-language neighborhood effects. Connectionist models have suggested 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 word neighbors, including *boor*, *door*, *goor*, *hoor*, *koor*, *moor*) than do English words with few Dutch neighbors (such as *bath* with no Dutch word neighbors). Van Heuven et al. interpreted this 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.'s (1998) findings, Lemhöfer et al. found many more commonalities among the groups than differences, with substantial overlap in reaction time (RT) patterns and in the set of significant predictors across the participant groups. In

particular, there were nearly no influences of the bilinguals' mother tongues 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 of English and the participants' L1 (see also Diependaele et al.'s, 2013, study 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 being of no relevance for explaining overall interpopulation differences. Without knowledge of the interpopulation differences, it is tempting to (wrongly) generalize theoretical relevance to practical importance. One goal of this article was to present the community of researchers with a large data resource that could facilitate the examination of reading of English both as L2 and L1 across diverse language backgrounds and samples.

Background Literature

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., 2023). In contrast to earlier studies that had focused on targeted L1–L2 experimental manipulations, these recent studies adopted a megastudy approach, collecting large-scale data from samples of L1 and L2 participants in a naturalistic text reading task. For example, in the Multilingual Eye-Movements Corpus (MECO), Siegelman et al. (2022) and Kuperman et al. (2023) 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 can 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 the section Predictors of Reading Comprehension). Thus, 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 suggested a greater degree of behavioral constancy within a reader exposed to different languages than prior literature had suggested (see reviews by Godfroid, 2020; 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. (2023) were intriguing, but also limited because the eye-tracking approach requires access to expensive special equipment that 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 (e.g., Brysbaert, 2019; Schönbrodt & Perugini, 2013; Vermeiren et al., 2022). Our study, called ENglish Reading Online (ENRO), is administered fully online and does not involve eye tracking, which opens up studies to many more diverse groups and to bigger sample sizes than does the eye-tracking approach.

Predictors of Reading Comprehension

Reading researchers have generally agreed that reading comprehension can be explained by the combination of two latent variables: word reading, that is, 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 included these two latent variables (e.g., the simple view of reading; Gough & Tunmer, 1986; Hoover & Gough, 1990) have explained 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 that 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 that, 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, that is, word reading and comprehension.

Both these constructs and the skills that they represent have at least two important interrelated 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, 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 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. Jeon and Yamashita (2022, an update of Jeon & Yamashita, 2014) recently published a relevant meta-analysis of the predictors of L2 reading comprehension in English and other languages. This meta-analysis showed strong positive correlations of L2 reading comprehension with, in decreasing order, L2 listening comprehension ($r = .81$), knowledge of L2 vocabulary ($r = .72$) and grammar ($r = .70$), L2 oral reading fluency ($r = .64$), knowledge of L2 morphology ($r = .64$), L2 phonological awareness ($r = .61$), knowledge of L2 orthography ($r = .59$), and L1 reading comprehension ($r = .48$). General cognitive resources (e.g., working memory and metacognition) only showed medium correlations with L2 reading comprehension (both around $r = .33$). Overall, these results suggested that L2 reading comprehension is more strongly correlated with L2 knowledge and skills than with L1 reading comprehension, leading Jeon and

Yamashita (2014) to conclude that "L2 reading comprehension is essentially determined by L2 language ability" (p. 189). Clearly, these meta-analyses are an incredibly rich source of information based on decades of rigorous research. Yet, their results were limited in important ways, namely that: (a) the estimates were based on different studies, often using different protocols and measures, (b) there was no information on how the different predictors correlated with one another, and (c) there was no comparison of L1 and L2 participants.

We set up ENRO 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 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 the variables explored in previous research. In particular, we chose to focus primarily on language and reading measures of 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 has provided information on reading comprehension, listening comprehension, reading rate, and multiple key component skills of English proficiency.

The Present Study

The ENRO study sought 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 article were two-fold: to introduce ENRO to reading researchers and to investigate similarities and differences in reading patterns in L1 and L2. We achieved the first goal 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 report can only shed partial light on the second goal. In this article, we confined ourselves to two outstanding questions motivated by our literature review: (a) Do L1 and L2 speakers differ on average, and if so, on which skills? (b) Do the interrelations of English skills vary between L1 and L2 readers, and if so, in what ways? We present three sets of analyses to shed light on these questions. First, we used regression models to estimate how much vari-

ance was explained by the L1–L2 distinction in the various ENRO measures. Second, we applied correlational and factor analyses to the data from the English L1 and L2 participants and compared how similar or divergent were the interrelations of various measures and constructs representing different facets of English proficiency in these two populations. Third, a partitioning-of-variance analysis asked to what extent skills and intersample differences predicted reading comprehension and rate as well as listening comprehension. This analysis quantified the relative role played by the L1 versus L2 contrast in reading and listening outcomes compared to a variety of cognitive, linguistic, and demographic variables. Following previous studies (e.g., Busby & Dahl, 2021; Dirix et al., 2020; Kuperman et al., 2023), we also probed 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.

We administered all assessments and questionnaires of the ENRO database 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 L1s. These samples included a large number of both L1 and L2 readers of English (3,853 and 3,485, respectively). The languages of instruction in the institutions where ENRO participants were enrolled demonstrated 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). The analysis of this variability can enable researchers to tease 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 that 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.

Method

Participants

The ENRO data included a total of 7,338 participants. These participants were recruited in 30 partner sites that 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 (https://www.prolific.co). The samples represented 19 countries and 16 unique source languages, defined as the language of instruction in the university where data were collected (which occasionally diverged from the language of the country or region[1]). Nine of the samples had English as the source language (four in Canada, one in New Zealand, one in the United Kingdom, two in the United States, and one sample of L1 English-speaking participants from the United States, the United Kingdom, and Canada recruited through Prolific). Although located in English-speaking countries, these sites included both L1 and L2 speakers of English.

The remaining 21 samples were recruited in countries and academic institutions where English is not an official or dominant language: These included 20 university-based samples and one sample of Dutch speakers recruited via Prolific. In these 21 samples, we asked partner sites to exclude, to the extent possible, individuals who had had an uncharacteristically intensive exposure to English—students who had lived in an English-speaking country for more than 6 months or were 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 the respective screening filters in the crowdsourcing samples. Thus, our samples were fairly homogeneous in educational status and age.

We requested participating sites to collect at least 100 participants; 26 of the 30 participating sites reached this sample size. We decided also to include in the ENRO data the four remaining sites ($n > 50$ in all samples) to avoid data loss. All the participants included in ENRO provided accuracy data from the central reading comprehension task. There were 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 in the Supporting Information online, we have listed the number of participants with valid data from each measure and provided more details about reasons for missing data. The number of missing values was generally small, with one exception: The listening comprehension task, which did not record responses between 12 February 2021 and 22 April 2021 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). Of the 7,338 participants on whose data we have reported in this paper, 4,875 participants completed all tests including listening comprehension ($n_{L1} = 2{,}615$; $n_{L2} = 2{,}260$).

Our analyses ranged from the individual level, in which a single participant was 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 L1 speakers of English as those who indicated in the language background survey that they had first started to learn to speak English before the age of 5 years, that is, the age around which formal schooling begins in many participating countries. Below we discuss implications of this definition for our results. For convenience, we labeled all other participants as L2 English speakers, even though for some of the participants English may have been their third or fourth language. We separately considered subsamples of L1 and L2 speakers at sites that showed a substantial representation of both types of speakers. All such samples were found in universities in Canada and the United States, and hence each of these sites included subgroups of both L1 and L2 speakers (see Table 1). A small percentage of participants (< 5%) did not self-designate as L1 in the United Kingdom 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 years; see Table 1). To avoid data loss, we imputed the majority L2 status to such participants instead of removing them from consideration. 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 were at a university with English as a source language, we assigned them to the L1 group; otherwise to the L2 group. In the project's Open Science Framework repository (https://osf.io/gzyqf), we have included information regarding both the participants' original and imputed L1–L2 status, yet for the purpose of our analyses,

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	CA	McGill U	L1	English	61	54	60
2	ca_mcg_english_L2	CA	McGill U	L2	English	51	0	47
3	ca_mcm_english_L1	CA	McMaster U	L1	English	1,895	1,752	1,743
4	ca_mcm_english_L2	CA	McMaster U	L2	English	303	0	280
5	ca_ua_english_L1	CA	U of Alberta	L1	English	271	246	263
6	ca_ua_english_L2	CA	U of Alberta	L2	English	99	0	99
7	ca_uo_english_L1	CA	U of Ottawa	L1	English	759	670	661
8	ca_uo_english_L2	CA	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

(*Continued*)

Table 1 (Continued)

Unit		Country	University	English status	Source language	*N*	*N* L1	*N* LISN
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
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

Note. *N* L1 = number of participants in the unit who reported learning English before the age of 5; *N* LISN (Lectures, Interviews and Spoken Narratives) = 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.

we used 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 report, we have not distinguished between L2 speakers of English enrolled as students in English-dominant countries/institutions (English as a second language) and L2 speakers of English in non-English-dominant sites (English as a foreign language); we have left this for future research as well (see Limitations section).

As we noted above, we defined each resulting sample of participants as a combination of the recruitment site, the source language, and the imputed L1–L2 status of English. We have referred 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 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, we imputed and labeled all the participants 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 in the Supporting Information online.

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

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. Descriptions of additional ENRO instruments (tests of component skills and motivation survey) are available in Appendix S3 in the Supporting Information online.

Reading Comprehension

The 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 and universities. 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 the participants' ability to determine central ideas of the text, summarize and synthesize its content, and analyze argumentation, word choice, and text structure. Appendix S4 in the Supporting Information online 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 participants' percent of correct responses for the 45 questions. Twelve of the 15 texts in the reading comprehension texts were also used in the L2 component of MECO (Kuperman et al., 2023). 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 Open Science Framework repository (https://osf.io/gzyqf) 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]). We disregarded values that were unrealistically long (possibly reflecting distraction or connectivity issues) or short (possibly reflecting a response after only partial reading, skimming, or skipping). We considered only reading rates in the interval between 89 and 804 wpm, that is, reading rates that were 3 times slower or faster than the estimated mean reading rate of a L1 reader (268 wpm; see Brysbaert, 2019, and Just & Carpenter, 1987; see also Kuperman et al., 2021). We then computed a measure of mean reading rate across texts for each participant.

Listening Comprehension

The test was an adaptation of the Lectures, Interviews and Spoken Narratives Test (Sommers et al., 2011; Tye-Murray et al., 2008), a listening comprehension test that consisted of five recorded audio passages based on narratives selected from the Rutgers University Oral History Archives of personal

descriptions of life experiences (Sommers et al., 2011). The passages were between 1 and 2 min long and were recorded by male and female professional actors with North American accents. ENRO used edited versions of five 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 for the 25 questions was a participant's score in the task that we labeled as listening comprehension.

Demographic and Language Background Questionnaire

All the participants completed the Brief Social and Language History Questionnaire that aims 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) that was originally adapted from various questionnaires used in the field of bilingualism, in particular, the Language Experience and Proficiency Questionnaire (Kaushanskaya et al., 2020) and the Language History Questionnaire 3.0 (Li et al., 2020).

The first part of the survey included questions about the 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 that each participant read and spoke. In each of these sections, the participants first listed the languages that they spoke/read (either by selecting them from a dropdown menu of languages or by typing them manually). Then, for languages the participants 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's (2021a) study, we used percent-based scales (e.g., "Indicate your current percentage use of all of the languages you speak, in each of the following environments").

Component Skills of English Reading Proficiency and a Motivation Questionnaire

Seven tests tapped into component skills of English reading proficiency: a grammaticality judgment task, a spelling recognition task, a vocabulary

knowledge task, an orthographic awareness task, a text segmentation task, the Lexical Test for Advanced Learners of English (LexTALE), and a lexical decision task. The eighth test tapped into how motivated our participants were to excel in the study. As we noted above, we chose the tests because of their theoretical relevance (i.e., to provide measures of word reading and listening comprehension as well as foundational language knowledge) and for practical considerations for online administration and study duration. Several of these component skills were equally relevant for listening comprehension (e.g., vocabulary knowledge, grammatical knowledge), but others were 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 we could therefore use them to determine ENRO's convergent validity. Other considerations behind test selection included maintaining backward compatibility with MECO L2 scores—the MECO used the same measures of vocabulary knowledge, spelling recognition, and LexTALE (see Appendix S3 for further details on the test stimuli and scoring, including references).

Procedure

We administered the study online using an in-house web-based data collection platform. The 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 research ethics board. The 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 task,
3. grammaticality judgment task,
4. listening comprehension task,
5. spelling recognition task,
6. vocabulary knowledge task,
7. motivation questionnaire,
8. orthographic awareness task,
9. text segmentation task,
10. LexTALE, and
11. lexical decision task.

The entire task battery typically took about 1.5 hours for the participants to complete; the time rarely exceeded 2 hours.

Table 2 Reliability estimates for ENglish Reading Online (ENRO) project measures

Measure	Reliability: Full sample	Reliability: L1	Reliability: L2
Reading comprehension	.85	.86	.85
Reading rate (Comprehension task texts)	.96	.95	.96
Listening comprehension	.80	.80	.80
Motivation	.81	.83	.78
Vocabulary	.93	.94	.92
Spelling	.81	.80	.80
Grammatical knowledge	.81	.69	.79
Lexical decision: Accuracy	.98	.98	.96
Lexical decision: Reaction time	.98	.98	.97
LexTALE: Accuracy	.90	.90	.86
LexTALE: Reaction time	.95	.91	.96
Orthographic awareness	.74	.76	.72

Note. For all measures but reading rate, we used a split-half procedure; we corrected estimates with the Spearman-Brown formula. For reading rate, we used the intraclass coefficient across the 15 texts.

Data Availability

The ENRO project has committed itself to the 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. The full code and data are available through the project's Open Science Framework repository (https://osf.io/gzyqf).

Results

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 better fitted an intraclass coefficient analysis (i.e., examining agreement in reading rates across the relatively small number of 15 texts). For split-half estimates, we computed Spearman-Brown corrected values (Spearman, 1910), which reflect reliability for the full sample of items in each test (rather than for the half of the items that were the bases for uncorrected correlation estimates). Table 2 shows that reliability was high for all

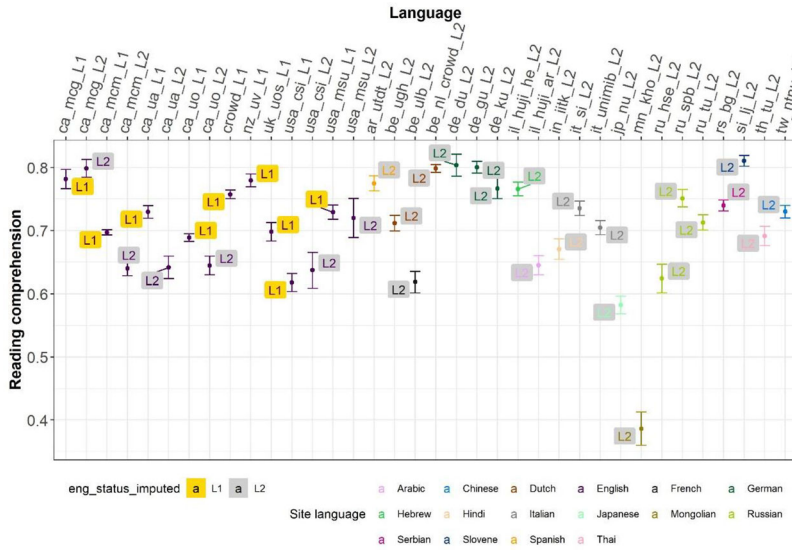

Figure 1 Reading comprehension scores (proportion correct) according to participants' imputed status as first (L1) or second language (L2) speakers. Error bars stand for ±1 standard error.

ENRO measures in the full sample of participants (all estimates ≥ .74) and very high for the main variables of interest: reading comprehension, reading rate, and listening comprehension (all estimates ≥ .80). Table 2 also provides the reliability estimates computed separately for the L1 and L2 participants grouped on the basis of our measure of imputed L1–L2 status. These estimates confirmed that the measures' sensitivity was at satisfactory levels for both L1 and L2 participants (all estimates ≥ .69).

Descriptive Statistics: Mean Reading Performance by Participant Unit

To obtain 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), we calculated the means and standard errors for all ENRO measures for each unit. For brevity, in the main text, we have provided descriptive plots for the 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 in the Supporting Information online. An auxiliary table with full descriptive information broken down by unit (e.g., number of participants, mean, standard deviation, median,

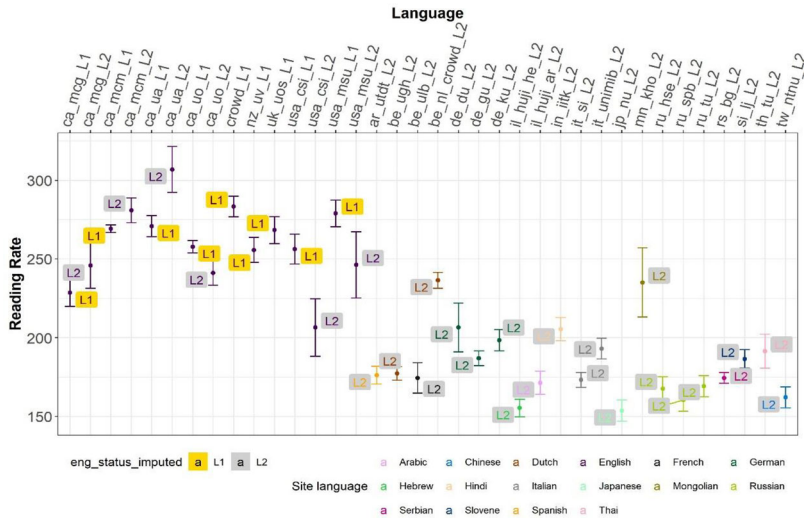

Figure 2 Reading rate (words per minute) according to participants' imputed status as first language (L1) or second language (L2) speakers. Error bars stand for ±1 standard error.

range, and standard error for each measure) is available through the project's Open Science Framework repository (https://osf.io/gzyqf).

A few noteworthy trends emerged from the results displayed in Figures 1 and 2. The difference in English reading comprehension accuracy among English-dominant (universities in Canada, New Zealand, the United Kingdom, and the United States) and nondominant sites appeared to be small. This replicated the finding reported by Kuperman et al. (2023). Yet, reading comprehension varied considerably between testing sites representing the same languages whether we considered the English L1 or L2 participants. It is worth noting that we observed similar patterns in listening comprehension performance (see Appendix S5). In contrast, just like in MECO L2 component (Kuperman et al., 2023), estimates of mean reading rate (Figure 2) demonstrated 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.

Comparison of L1 and L2 Readers' Mean Performance Across Tasks

This analysis investigated how much variance the L1–L2 distinction explained in participants' performance in all experimental tasks of ENRO tapping into

the English language processing. This information was useful for understanding which component skills or outcomes of reading showed greater or lesser overlap between these groups of participants. Table 3 shows the means and standard deviations of the L1 and L2 participants 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 included only the participants without missing values in any of the dependent variables, to ensure that estimates were based on the same set of participants for all measures ($N = 5{,}023$).

This analysis quantified 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, English L1 and L2 participants 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 participants' 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 indicated where the coarse-grained differences between the groups of L1 and L2 participants lay and paved the way for more in-depth analyses.

Interrelations of English Skills Among L1 and L2 Participants

We next turn to analyses quantifying the interrelations of ENRO measures, and, primarily, the extent to which these interrelations varied between the L1 and L2 participants. Thus, complementary to the estimates of main effects in the last section (i.e., the overall differences in mean performance between the L1 and L2 participants), analyses in this section tapped 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 correlations among test scores, computed separately for the L1 and L2 participants. In Appendix S6 in the Supporting Information online, we have further provided estimates for the correlations after correction for attenuation given the tests' reliability, as well

Table 3 Means and standard deviations (in parentheses) of L1 and L2 participants, 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	L2 participants	Cohen's *d*	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)	267.30 (106.35)	198.54 (94.31)	0.68	10%
Text segmentation	41.31 (13.14)	32.92 (13.58)	0.63	9%
LexTALE: Reaction time	968.96 (283.35)	1,200.55 (475.23)	0.60	8%
Lexical decision: Reaction time	720.63 (114.43)	802.76 (158.63)	0.60	8%
Spelling	0.86 (0.09)	0.81 (0.11)	0.48	5%
Listening Comprehension	0.64 (0.18)	0.6 (0.20)	0.22	1%
Orthographic awareness	0.89 (0.10)	0.87 (0.09)	0.20	1%
Reading comprehension	0.73 (0.15)	0.73 (0.15)	0.00	< 1%
Motivation	3.59 (0.60)	3.57 (0.55)	−0.03	< 1%

Note. LexTALE = Lexical Test for Advanced Learners of English.

Table 4 Uncorrected Pearson correlations between ENglish Reading Online (ENRO) project measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. read_comp_score	–	.32	.69	.65	.45	.46	.55	.27	.62	.31	.42	.47	−.30	.11	.10	−.08	−.10
2. motivation	.34	–	.36	.23	.13	.16	.23	.14	.27	.14	.15	.18	−13	.03	.04	−.07	−.04
3. lisn_score	.68	.33	–	.53	.38	.40	.51	.26	.54	.27	.33	.39	−.22	.07	.07	−05	−.08
4. vocab_score	.59	.29	.55	–	.49	.53	.59	.22	.63	.29	.45	.48	−.11	.20	.18	−.16	−.15
5. spell_score	.48	.23	.42	.61	–	.36	.45	.10	.47	.14	.36	.40	−.01	.19	.22	−.06	−.09
6. grammar_score	.49	.24	.61	.53	.38	–	.47	.08	.46	.13	.34	.40	−.12	.22	.19	−.15	−.14
7. lextale_score	.52	.27	.56	.62	.53	.59	–	.21	.77	.26	.42	.46	−.10	.19	.17	−.15	−.13
8. lextale_rt	.21	.08	.13	.13	.12	−.05	.13	–	.27	.63	.18	.03	−.31	−.03	−.06	.00	−.02
9. ld_score	.61	.29	.62	.63	.55	.63	.77	.13	–	.44	.45	.46	−.14	.15	.14	−.14	−.13
10. ld_rt	.32	.14	.20	.22	.19	.07	.18	.64	.38	–	.21	−.04	−.31	−.01	−.01	−.01	−.02
11. ortho_score	.40	.15	.29	.34	.37	.26	.31	.12	.41	.21	–	.34	−.10	.09	.09	−.06	−.08
12. segment_score	.49	.24	.47	.43	.44	.46	.51	−.08	.57	−.01	.35	–	.00	.12	.13	−.13	−.15
13. rate_mean	−.26	−.08	−.17	.03	.00	−.01	.03	−.21	−.06	−.27	−.13	−.01	–	.05	.08	−.03	−.03
14. eng_speech_proficiency	.27	.19	.40	.41	.32	.55	.45	−.17	.42	−.08	.13	.32	.06	–	.77	−.16	−.11
15. eng_read_proficiency	.27	.21	.37	.39	.35	.50	.43	−.18	.41	−.10	.15	.33	.06	.82	–	−.10	−.11
16. eng_speech_age	−.10	.00	−.09	−.08	−.10	−.06	−.09	.06	−.09	.03	−.06	−.15	−.01	−.15	−.12	–	.42
17. eng_read_age	−.07	−.01	−.08	−.04	−.08	−.07	−.06	.03	−.06	.02	−.07	−.11	−.04	−.13	−.13	.61	–

Note. Values above the diagonal show correlations among L1 participants (maximum $n = 3{,}853$) and below the diagonal among L2 participants (maximum $n = 3{,}485$). read_comp_score = reading comprehension; lisn_score = listening comprehension; vocab_score = vocabulary knowledge; spell_score = spelling; grammar_score = grammatical knowledge; lextale_score = Lexical Test for Advanced Learners of English, accuracy; lextale_rt = Lexical Test for Advanced Learners of English, mean reaction time; ld_score = lexical decision, accuracy; ld_rt = lexical decision, mean reaction time; 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.

as correlation estimates computed over the full sample (i.e., including both the L1 and L2 participants).

Table 4 points to a good convergence between the ENRO data and In'nami et al.'s (2022) and Jeon and Yamashita's (2014, 2022) meta-analyses of reading and listening comprehension studies. 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 measured through lexical decision and LexTALE tests), of grammar knowledge, and of orthographic awareness. To further compare previous meta-analytic estimates to the ENRO data, in Appendix S7 in the Supporting Information online, we present estimates of predictors of reading and listening comprehension in L2 participants from Jeon and Yamashita's (2014, 2022) and In'nami et al.'s (2022) meta-analyses, alongside ENRO estimates for both L1 and L2 participants. Overall, the convergence of estimates among the L2 participants gave additional credibility to both the measures overlapping with estimates in the meta-analyses and with the estimates for the additional variables that we introduced: motivation to excel, subjective speech and reading proficiency, and skill tests. Importantly, the similarity in the relative strength of correlations also spanned the range of the correlations for the L1 ENRO participants, who demonstrated a strong role for grammar, vocabulary, and orthographic knowledge as predictors of their reading and listening comprehension similar to their role as predictors for the L2 ENRO participants and as predictors in the meta-analytic studies.

Next, and central to our theoretical question, we compared the correlations computed separately for the 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 applying a Bonferroni correction for multiple comparisons. Table 5 presents these results. Notably, in most cases, the L1 and L2 participants showed a similar magnitude of correlations among the ENRO tests. In fact, of 136 pairwise comparisons, only 12 exceeded an absolute value of .20, and 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 were due to the lack of variability in responses among the L1 English participants (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 among

Table 5 Comparison of the magnitudes of correlations among L1 and L2 participants

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. read_comp_score	–	−.02	< .01	**.06**	−.03	−.04	.02	.06	< .01	−.01	.02	−.02	−.03	**−.16**	**−.16**	.02	−.04
2. motivation	.337	–	.03	−.06	**−.10**	−.08	−.04	.06	−.02	< .01	−.01	−.06	−.06	**−.15**	**−.17**	−.07	−.03
3. lisn_score	.483	.205	–	−.02	−.04	**−.21**	−.05	**.13**	**−.08**	.07	.05	**−.09**	−.06	**−.33**	**−.30**	.03	.01
4. vocab_score	< .001	.006	.295	–	−.12	< .01	−.03	**.09**	< .01	.07	**.11**	.05	**−.14**	**−.21**	**−.22**	−08	**−.10**
5. spell_score	.102	< .001	.077	< .001	–	−.02	**−.08**	−.02	**−.08**	−.05	−.02	−.04	−.01	**−.13**	**−.13**	.04	−.01
6. grammar_score	.100	< .001	< .001	1.000	.326	–	**−.12**	**.13**	**−.17**	.06	**.08**	−.06	**−.11**	**−.33**	**−.32**	**−.09**	−.07
7. lextale_score	.073	.069	.009	.044	< .001	< .001	–	.09	< .01	**.09**	**.11**	−.05	**−.13**	**−.26**	**−.26**	−.07	−.07
8. lextale_rt	.007	.010	< .001	< .001	.389	< .001	< .001	–	.14	−.01	.05	.05	**−.10**	**.14**	**.11**	−.06	−05
9. ld_score	.493	.355	< .001	1.000	< .001	< .001	1.000	< .001	–	.07	.04	**−.11**	−.08	**−.27**	**−.27**	−.05	−.07
10. ld_rt	.636	1.000	.006	.002	.029	.010	< .001	.475	.002	–	< .01	−.03	−.04	.07	.09	−.04	−.03
11. ortho_score	.306	1.000	.101	< .001	.624	< .001	< .001	.009	.037	1.000	–	−.01	.02	−.04	−.07	.01	−.01
12. segment_score	.274	.008	< .001	.008	.041	.002	.006	.035	< .001	.207	.633	–	.01	**−.20**	**−.20**	.02	−.03
13. rate_mean	.069	.034	.058	< .001	.676	< .001	< .001	< .001	< .001	.068	.205	.680	–	−.01	.02	−.01	.01
14. eng_speech_proficiency	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	.005	.103	< .001	.691	–	**−.06**	−.01	.01
15. eng_read_proficiency	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	.017	< .001	.438	< .001	–	.01	.02
16. eng_speech_age	.415	.005	.157	.001	.104	< .001	.014	.016	.041	.107	1.000	.416	.427	.679	.427	–	**−.20**
17. eng_read_age	.236	.239	1.000	< .001	.693	.006	.006	.051	.006	.118	.694	.115	.700	.426	.426	< .001	–

Note. Values above the diagonal show correlation differences ($r_{L1} - r_{L2}$). Values below the diagonal are p values for each difference based on a Fisher's r-to-z transformation. Correlation differences significant after Bonferroni correction ($\alpha = 5\%/136$) are shown in bold. read_comp_score = reading comprehension; lisn_score = listening comprehension; vocab_score = vocabulary knowledge; spell_score = spelling; grammar_score = grammatical knowledge; lextale_score = Lexical Test for Advanced Learners of English, accuracy; lextale_rt = Lexical Test for Advanced Learners of English, mean reaction time; ld_score = lexical decision, accuracy; ld_rt = lexical decision, mean reaction time; 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.

experimental tasks that showed a L1–L2 difference larger than |.20| was between grammatical knowledge and listening comprehension ($r_{L1} = .40$; $r_{L2} = .61$). We concluded that interrelations of ENRO tests were generally similar in magnitude among the L1 and L2 participants. Still, some multiple correlation differences reached statistical significance: Even after excluding correlations with self-report measures, 25/78 (32%) correlations were significant after applying a Bonferroni correction. Most of these statistically significant L1–L2 differences were practically small and involved either grammatical knowledge (seven correlations) or LexTALE (six in both accuracy and six in RT), suggesting that component skills of English proficiency that most differentiated the L1 and L2 participants in performance (see Table 3) were also the ones that showed the greatest difference in predictive power between the L1 and L2 groups. More broadly, this result pointed again to the distinction between statistical, theoretical, and practical significance of L1–L2 comparisons.

Factor Analysis

The correlational analysis tapped into interrelations of variables defined at the single-task level and examined whether and how these correlations varied between the L1 and L2 participants. A logical next step was to investigate how the variables loaded onto latent factors and whether this grouping was different between the L1 and L2 participants. Exploratory factor analysis provided an initial answer to this question, testing how the various tests grouped together and how similar the solution was for the L1 and L2 participants (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, 2023) indicated that three factors in both the L1 and L2 participant groups accounted for more than half the variance (51% for L1 and 54% for L2). Table 6 presents the results of the two factor analyses that we conducted on the two samples.

Among both the L1 and L2 participants, multiple accuracy-based tests of English component skills grouped into Factor 1. These included accuracy from lexical decision, spelling, vocabulary, grammar, and text segmentation (orthographic awareness loaded onto this factor for the L1 but not for the L2 participants). Also reflecting similarity in the L1 and L2 groups, the RT measures in lexical decision and LexTALE tasks (after logarithmic transformation) grouped into Factor 2 that reflected response slowness. The two comprehension scores (listening and reading) loaded onto Factor 3, a factor which correlated strongly ($r_{L1} = .64$; $r_{L2} = .55$) with Factor 1, the Proficiency Factor. Finally, in both the L1 and L2 participants, reading rate loaded onto

Table 6 Exploratory factor analyses for L1 and L2 participants

	L1 participants			L2 participants		
Measure	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%

Note. Variable loadings (absolute values higher than 0.30 are presented) and cumulative variance explained. Additional information can be accessed via the project's Open Science Framework repository (https://osf.io/gzyqf). read_comp_score = reading comprehension; lisn_score = listening comprehension; vocab_score = vocabulary knowledge; spell_score = spelling; grammar_score = grammatical knowledge; lextale_score = Lexical Test for Advanced Learners of English, accuracy; lextale_rt = Lexical Test for Advanced Learners of English, 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).

Factor 3. Fast readers in both samples had lower comprehension scores. Together, the results of the factor analysis suggested that, although there were subtle differences between the L1 and the L2 participants (e.g., loading of comprehension scores and reading rate on Factor 1 for the L2 but not the L1 participants; loading of reading rate on Factor 2 for the L1 but not the L2 participants), the similarities in how variables of English proficiency grouped to latent factors in the L1 and the L2 participants displayed substantial overlap.

What Explains Variance in Reading and Listening Comprehension?

Analyses so far estimated the overall differences between the L1 and L2 participants in their English-language performance and the similarities and differences in interrelations of test scores (and their loadings on latent factors) in the two participant groups. We additionally tackled a related theoretical issue:

What are the sources of interindividual variability in English reading and listening comprehension? Unlike comparative quantification of the group effect of the L1–L2 distinction presented above, this analysis enabled us to examine the possible L1–L2 similarities and differences when we controlled for other variables that explained variability in the English performance (specifically, the English component skills). The analysis also estimated the predictive value of the L1–L2 difference relative to other sources of variance.

In all analyses in this section, we treated 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 we decomposed the variability in the three outcome variables into the variability explained by four components:

- English component skills,
- the L1–L2 distinction, that is, whether a participant was or was not a L1 English speaker,
- the L1 of the participant (among L2 speakers), and
- the intersample 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 where we added a group of predictors at each step to examine the additional unique amount of variance explained by the predictors added (see Kuperman et al., 2023, for a similar approach).

We added predictors to the models in four steps. At Step 1, we added variables that reflected 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 model, Hoover & Gough, 1990). Variance explained at this step reflected the overall role of component skills, measured 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 reflected the variance explained by the distinction between L1 and L2 participants while we controlled for the effect of English component skills. At Step 3, we added a set of dummy-coded variables coded for the participants' L1s other than English. Each variable in the set represented one L1 spoken by the participants in our sample, with L1 speakers of a given language coded as 1 and L2 speakers of that language coded as 0.[3] This step reflected added variance associated with differences between the L2 English participants of

different L1s (i.e., the effect of the participants' L1 on their L2 English performance), beyond the L1–L2 effect and the impact of component skills. At Step 4, the final step, we added a categorical variable reflecting differences between sites within a country (implemented as a series of dummy-coded variables). Thus, for example, this variable distinguished between the two different English-dominant sites in the United States (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 reflected additional variables that we expected to vary across sites within a country and language (e.g., differences in educational background, English entry requirements, socioeconomic status, etc.). For comparability, at all steps, we included only the participants with complete data ($N = 5{,}023$). Figure 3 presents the outcome of this analysis.

In analyses of reading and listening comprehension accuracy (two left columns in Figure 3), the vast majority of variance was explained at Step 1 by the English component skills of listening and reading (the model predicting reading comprehension 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 amounted 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 ($\Delta R^2 = 2.1\%$ for both reading and listening comprehension, which amounted 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., the participants' L1 and the site within a country and the L1) led to little improvements in variance explained: all $\Delta R^2 \leq 1.6\%$, which constituted 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 the 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 3). First, the total variance explained by the variables at Steps 1–4 was lower ($R^2 = 26.5\%$ in total). Most likely, this reduction in R^2 was related to the lesser impact of English component skills ($R^2 = 16.7\%$ at Step 1 for reading rate, less than half of the impact of English component skills 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, the L1–L2 distinction ($\Delta R^2 = 3.2\%$, accounting for 12.0%

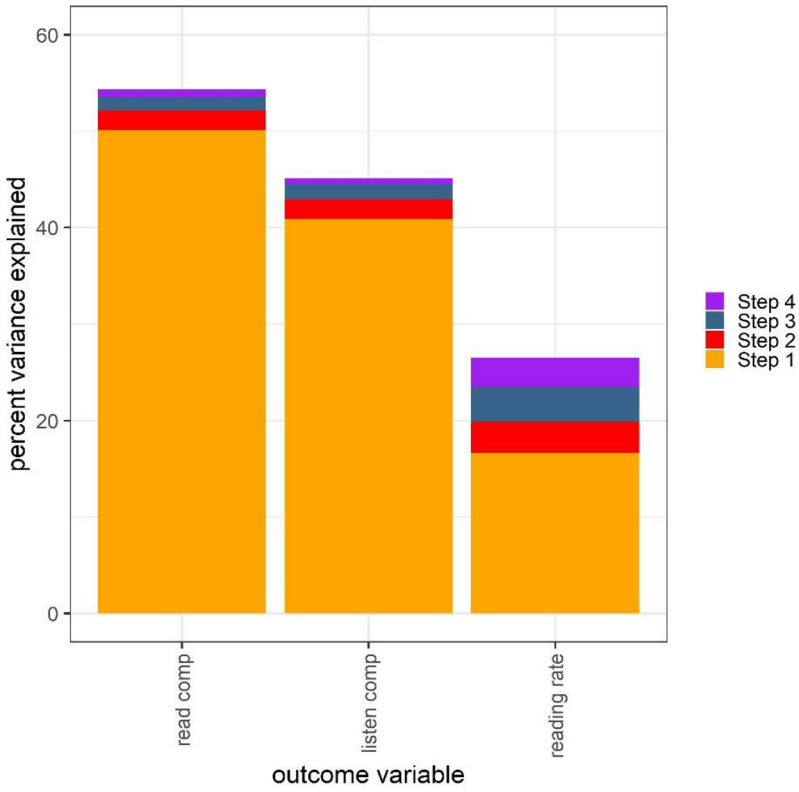

Figure 3 Stepwise partitioning of variance in English reading comprehension (read comp), listening comprehension (listen comp) and reading rate. Step 1: component skills of English; Step 2: differences between L1 and L2 speakers; Step 3: differences between L1s (other than English); Step 4: intersite variability within a country and L1.

of total variance explained) and L1 background ($\Delta R^2 = 3.6\%$, accounting for 13.5% of total variance explained) explained larger relative portions of the variance. Cross-site differences within countries and native languages (Step 4) accounted for additional variance at a level comparable to that explained by Steps 2 and 3: $\Delta R^2 = 3.1\%$, that is, 11.6% of total variance.

We note that the order in which we entered the variables into analyses 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 we had controlled for this variance in Step 1). However, alternative orders could be used to examine related questions. We

explored one such alternative ordering of variables, putting the L1–L2 distinction before component skills, in Appendix S8 in the Supporting Information online. Crucially, this analysis still replicated the key finding above for 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).

Discussion

Research into L2 reading has 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., 2023, and the Introduction section). The first goal of this article was to introduce to the research community ENRO, a new data source that fits this description. The ENRO database contains data from 7,338 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 L1 or as a language that they had acquired before the age of 5 years: Under this criterion, discussed in detail below, we considered them to be L1 speakers of English (see Methods section for the imputation procedure). Most other participants were advanced learners of English who had passed English language examinations to be accepted into 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). We have made the trial-level data from all tasks publicly available. Furthermore, detailed questionnaires provide rich data on language background and use as well as demographic characteristics of the participants (Gullifer & Titone, 2020). As envisioned by the study design, the ENRO data make possible both big picture exploratory studies of L1 and L2 proficiency and targeted investigations of language and reading behavior driven by specific theoretical questions or focusing on subsets of language backgrounds, participants, or items. This first report confined itself to description and

methodological validation of the data collected and focused on only a few big-picture questions. The questions, broached in detail in the Introduction, were: (a) How similar or different are language and reading behaviors of L1 and L2 readers of English? (b) What predictors explain variance in measurable outcomes of English reading and listening comprehension?

Methodological Foundations: Reliability and Validity of English Reading Online (ENRO) Measures

ENRO assessments in the full sample of participants showed high reliability in all outcome measures, and reliability estimates were especially high for the reading and listening tasks. Similarly, we observed high reliability in the L1 and L2 participants when we considered the groups separately. This suggests high stability and utility of our data for investigation of individual differences. Furthermore, the validity of the ENRO data was supported by the correlations between ENRO measures and these measures' compatibility with the results of recent meta-analyses of reading and listening comprehension (In'nami et al., 2022; Jeon & Yamashita, 2014, 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 our primary study (see Appendix S7). Together, these analyses show that our data source both provides high-quality data and can obviate the shortcomings present in current studies (shortcomings that also have placed limits on the extant meta-analyses), for example, heterogeneity of studies and samples, lack of information for the correlations between the different predictors, and lack of direct comparison of L2 with L1 participants.

How Does Language and Reading Behavior Compare Across English L1 and L2 Participants?

This question is central for our article and stems from the bias that existing studies in the field have toward emphasizing differences rather than similarities between L1 and L2 speakers of a language under examination. As we argued in the Introduction, the bias is at least partially grounded in the null hypothesis-testing logic that is designed to detect how much 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 versus difference between L1 and L2 populations. The goal was to provide empirical grounding to the question of

how fruitful it is to adhere to the L1 versus L2 binary distinction rather than to consider English language and reading proficiency across a continuum of skill, ability, and experiences that span L1 and L2 speakers alike (Diependaele et al., 2013; Gullifer et al., 2021).

Our data showed overwhelming evidence for similarities, rather than differences, between the L1 and L2 participants in their reading behavior and relative contributions of component skills of English reading to this behavior. Perhaps the most telling finding in this regard was the very small difference in reading and listening comprehension levels between the L1 and L2 participants (Cohen's $d = 0.00$ and 0.22, respectively; see Table 3). Not all component skills of English comprehension showed similarly small differences in L1 versus 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 ($d = 0.92$; 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 for the L1 and for the L2 participant 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 observation that relative roles played by component skills in reading comprehension found further support in correlational and factor analyses that examined interrelations of predictors and outcomes of comprehension tasks. The magnitudes of the correlations among ENRO test scores were highly similar between the 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 exert a similar influence on English reading and listening comprehension and reading rate in the 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 that 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 (a) untimed component skills of English proficiency (e.g., vocabulary, grammar, orthography, and spelling knowledge, and

accuracy in lexical decision tasks), (b) timed responses (lexical decision RTs), and (c) comprehension scores (reading and listening). In both exploratory factor solutions, one of the factors on which reading rate loaded was comprehension. Faster readers showed lower comprehension levels, that is, 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. For example, only in the L2 sample did listening and reading comprehension and reading rate load 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; for example, 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 interrelations of component skills and outcomes of English reading comprehension and their attribution to latent constructs.

Last, the partitioning-of-variance analysis, showed that the contribution of the L1–L2 contrast to explaining variance in the main outcomes of the 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 the participants attending different universities within a country were comparable to the differences associated with the L1–L2 contrast. In contrast, we traced the vast majority of explained variance in reading and listening comprehension (over 90%) and reading rate (63%) back to 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 the specific L1s reported by the L2 speakers of English in our dataset, or the binary L1–L2 distinction) were 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., 2023; Nisbet et al., 2022). This suggests that a fruitful approach for further studies of university-level advanced learners of English would therefore concentrate on the shared nature of language and reading acquisition and knowledge rather than on the demonstrably small differences.

What Explains Variance in Reading and Listening Comprehension and Reading Rate?

As we 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 predicted these outcomes. They highlighted the strong role of the same higher-order skill set as indicated in meta-analyses of L2 comprehension (In'nami et al., 2022; Jeon & Yamashita, 2014, 2022), that is, 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 language speakers coordinate and rely on component skills 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 of reading rate, on the other hand. These differences—emphasized earlier by Busby and Dahl (2021), Dirix et al. (2020), and Kuperman et al. (2023), among others—emerged in all analyses that we 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 the English L1 and L2 participants (Cohen's $d = 0.68$). We did not observe this distinction in either the 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 contrasted 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 the 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., 2023, 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 that we considered were statistically matched in their reading and listening comprehension performance (see also Vander Beken et al., 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, given that reading comprehension and reading speed are influenced by different developmental variables and rely on largely different skills and abilities. As Kuperman et al. (2023) argued, on the basis of 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 nativelike performance in the quality of comprehension among advanced learners of English does not come with nativelike reading speed. Yet speed is of obvious importance for workplace and academic environments that often place 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 attention from the current focus on reading and listening comprehension toward 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, but causes of that variability and even its direction—slower is better—are not yet entirely understood.

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 participants were mostly advanced university-level speakers of English, often with early and intensive exposure to English. Also, for simplicity, we glossed over the distinction between L2 speakers studying in English-dominant versus non-English-dominant institutions (e.g., English as a second language vs. English as a foreign language), leaving the investigation of this distinction 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 our results. First, we defined L2 speakers of English as those who had acquired English at or after the age of 5 years, which is a common age for starting formal schooling in many participating countries. Although adopted by some researchers as an operationalization of the L1–L2 distinction, this threshold is not universal. For example, 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 versus L2 speakers of English—to the degree that a person exposed to English from, for example, the age of 1 year differs from a person exposed to English from the age of 5 years. 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. We labeled all the participants in L2 sites as L2 speakers of English (regardless of self-reports), and in L1 sites where the L1–L2 distinction in the sample was too small (fewer than 5%), we relabeled the self-reporting L2 participants as L1s. The availability of ENRO data and code, including, in particular, the rich language background data collected, make it possible for researchers to validate the present findings against alternative and more fine-grained definitions of L1 or L2 speakers of English.

Further limitations relate to design choices that we had to make. The web-based nature of the study and time constraints led us to exclude from the test 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 L1 tests (for L2 speakers of English), which would allow comparisons of L2 reading rate and comprehension to L1 reading rate and comprehension. We did not include this aspect because it would be 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 that can focus, for example, on within-L1 analyses of L1–L2 reading comprehension and rate in specific languages of interest.

The findings of our 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 mask 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 examined only English as a target language, a language that is already massively overrepresented in L2 research. 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, our analysis focused on mean individual performance in each task, but we have also made available trial-level data for all tasks (where applicable). This reporting would make possible an in-depth investigation of, for example, lexical decision latencies and accuracy as a function of the participants' 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 the 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 L2 English have on L2 learning and proficiency (e.g., Schepens et al., 2013). We note that, in our population, this influence was likely to have been minor given the small group differences between the L1 and L2 participants that we 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.

Conclusion

This article introduced 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 participants' 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 the demonstrably high reliability of the tests contribute to the usefulness of the data for both bird's-eye-view comparisons of large groups of participants and the study of individual differences. We conducted analyses that addressed questions often posited as being central in 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 inquiry into reading behavior.

Final revised version accepted 4 April 2023

Open Research Badges

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data are available at https://doi.org/10.17605/osf.io/gzyqf.

Notes

1 Two noteworthy cases are McGill University, where we assigned English as a source language because it is an English-language institution in the predominantly French-speaking Canadian province of Quebec, and the Université libre de Bruxelles, where we assigned French as a source language given that it is a French-language institution in Belgium (a Dutch–French bilingual country). 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 the segmentation task was based on one trial only, we could not obtain reliability estimates for it.

3 To be consistent with our definition of L1–L2 status above, we defined a L1 speaker of a given language as someone who had learned the language before the age of 5 years. The participants could have multiple languages coded as their L1s. For simplicity, in this analysis, we included only languages that met our criterion of a L1 in 10 participants or more. The final list included 34 dummy-coded variables according to the languages chosen/entered by the participants (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 constituted a significant improvement in model fit (all $p < .05$). We expected this given ENRO's large sample size.

5 The factor analysis reported in Table 6 revealed 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 resulted from the cutoff that we used to flag strong loadings ($> |0.30|$). In this particular case, for instance, the same positive loading of reading rate onto Factor 1 existed in the L1s, only it was estimated at a subthreshold value of 0.23 (compared to 0.35; see the code on the project's Open Science Framework repository for full factor analysis output). Confirmatory factor analysis and structural equation modelling would be better techniques to use for answering 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*, *6*, 41–50. 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*, *26*(12), 1153–1170. 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*, Article 104047. https://doi.org/10.1016/j.jml.2019.104047

Brysbaert, M., & Rastle, K. (2021). *Conceptual and historical issues in psychology*. Pearson Higher Education.

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), 441–458. 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), 36–61. https://njes-journal.com/articles/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), Article 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*, *26*(1), 112–124. 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). Routledge. http://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*. 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* (3rd ed.). 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., Klein D., Baum, S., Phillips, N., & 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*, Article 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 E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 235–284). John Benjamins. 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 E. H. Jeon & Y. In'nami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 29–86). John Benjamins. https://doi.org/10.1075/bpa.13.03jeo

Just, M.A., & Carpenter, P. A. (1987). *Speedreading: The psychology of reading and language comprehension*. 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., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., Da Fonseca, S. M., Drix, N., Duyck, W., Fella, A., Frost, R., Gattei, C. A., Kalaitzi, A., Lõo, K., Marelli, M., Nisbet, K., … & Usal, K. A. (2023). Text reading in English as a second language: Evidence from the Multilingual Eye-Movements Corpus. *Studies in Second Language Acquisition*, *45*(1), 3–37. 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

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., Compton, D. L., Cho, E., & Patton, S., III. (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* (pp. 227–247). Blackwell. 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

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research* (R package Version 2.3.3) [Computer software]. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych

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* (pp. 199–230). De Gruyter Mouton. 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*. Routledge. https://doi.org/10.4324/9780203851357

Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H. D., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Da Fonseca, S. M., Dirix, N., Duyck, W., Fella, A., Frost, R., Gattei C. A., Kalaitzi, A., Kwon, N., Lõo, K., Marelli, M., … & Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods*, *54*, 2843–2863. 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. https://doi.org/10.1097/AUD.0b013e3182234cf6

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, *3*(3), 271–295.

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://doi.org/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*, *6*, Article 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*. Advance online publication. 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*, Article 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:

Accessible Summary

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.