



OPEN ACCESS

EDITED BY

Marijan Palmovic,
University of Zagreb, Croatia

REVIEWED BY

Kristina Cergol,
University of Zagreb, Croatia
Ying Zhao,
Inner Mongolia University of Science and
Technology, China

*CORRESPONDENCE

Rurik Tywoniw
✉ rtwywoniw@illinois.edu

RECEIVED 01 March 2023

ACCEPTED 15 May 2023

PUBLISHED 29 June 2023

CITATION

Tywoniw R (2023) Compensatory effects of individual differences, language proficiency, and reading behavior: an eye-tracking study of second language reading assessment. *Front. Commun.* 8:1176986. doi: 10.3389/fcomm.2023.1176986

COPYRIGHT

© 2023 Tywoniw. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Compensatory effects of individual differences, language proficiency, and reading behavior: an eye-tracking study of second language reading assessment

Rurik Tywoniw*

Department of Linguistics, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Reading in a second language (L2) is a complex process that incorporates linguistic knowledge and literacy abilities, as well as strategic competence to approach different types of reading tasks depending on reading goals. However, much of the previous research was limited to correlational studies and focused on the relative contribution of broad categories of L2 proficiency and first-language (L1) literacy to L2 reading comprehension. However, investigations into L2 reading performance can benefit from advances in real-time, concurrent data collection methodologies such as eye-tracking. This study utilized eye-tracking methods to examine L2 reading comprehension of 102 readers across three different reading tasks [Cloze reading, Multiple-choice (MC) quiz, and reading-to-summarize], comparing the comprehension scores to L2 proficiency, individual differences (reasoning, working memory, motivation) and reading behavior (eye-tracking metrics related to attention to reading texts and tasks, length of fixations). Results indicate that the score on each task could be modeled each using a different mix of predictors, with the cloze task being most strongly predicted and the MC task being least predicted. The Summary task was in-between, but with a highly interpretable model. Interactions between fixation duration and cognitive abilities were found, showing how efficient fixation is generally important for comprehension, but the impact can be compensated for with motivation and reasoning ability.

KEYWORDS

second language reading, language assessment, eye-tracking, English for academic purposes (EAP), language learning

Introduction

For multilingual readers and language learners, reading comprehension ability has been conceptualized as a product of language proficiency: learners reach a threshold of reading ability and can then transfer first language (L1) literacy skills (including comprehension monitoring, activating strategies, and integration of information across pieces of texts) into their second language (L2) reading (Koda, 1988, 1990). Features of reading comprehension that are not related to linguistic proficiency are often overlooked for multilingual readers. However, for many academic language learners in the modern era, advanced reading skills many develop uniquely for an L2 which is the primary language of academic engagement. As such, it remains unclear how features of reading comprehension processes which play a role in monolingual readers' comprehension, such as real-time reading behavior and individual differences, contribute to reading comprehension for multilingual readers. This

lack of understanding poses a threat to L2 reading assessment validity. [Bachman and Palmer \(1996\)](#) in their test-authenticity argument, state that use of a language test is justified when we can “demonstrate that performance on language tests corresponds to language use in specific domains other than the language test itself” (p. 23). To better understand the factors which influence reading comprehension performance for multilingual academic readers, it is necessary to compare factors of language ability, individual differences, and real-time reading, as well as comparing these factors’ influence on performance on varied reading tasks which may elicit different skills and abilities.

In this study, three measures of reading comprehension were analyzed: multiple-choice questions (MC), cloze tasks, and summary tasks. Completion of these tasks was analyzed under the lens of text-reading behavior. Task differences were examined using eye-movement behavior (eye-tracking) variables which were compared with score (described with more specificity in the methods section). Scores were predicted with statistical modeling using eye-movement metrics, L2 proficiency and individual difference variables: reading speed, working memory, reasoning, and motivation. This research will help the field of reading comprehension assessment further understand the cognitive and construct validity of these assessment tasks. Additionally, this research will shed light on how the influences on reading ability (individual differences, language proficiency, and real-time reading behavior) interact with each other and can be used to compensate for weaknesses.

Literature review

L2 reading and reading assessment

The validity of L2 reading tests hinges on how well tests target different aspects of the reading process. Models of reading often include both lower-order and higher-order skills. Key aspects of lower-level reading processes are grapho-phonemic processing, morphological awareness, word recognition, and syntactic parsing, with each lower-level process facilitating the recognition of words on the page ([Perfetti, 2007](#)). Much of the lower-order skills in L2 reading are developed alongside general L2 proficiency. Higher-level processing is seen as having two levels ([Kintsch, 1998](#); [Grabe, 2009](#)): a text base comprehension level, where a reader creates a model of ideas and propositional content found in a text, and a situation model level, where the overall meaning of a text is constructed by the reader through connecting propositions and relating content to background knowledge and reading context. L2 research has been more agnostic regarding the development of higher-order skills, believing much of this to be the recipient of L1 literacy transfer ([Koda, 1988](#)).

In general, L2 reading scholars have acknowledged that not every predictor of successful comprehension needs to be activated at once during reading. Early conceptions of this phenomenon considered L2 reading to be broken down into coarse categories of skills: L1 literacy and L2 language proficiency, and deficits in one category could be compensated for with strengths in the other ([Bernhardt, 2005](#)). This view was expanded beyond the broad categories of L1 literacy and L2 language ability to include

other potential compensatory strengths such as reading strategy knowledge and background knowledge ([McNeil, 2011, 2012](#)) in line with [Stanovich’s \(1980\)](#) postulation that individuals will rely on multiple top-down and bottom-up resources as needed to achieve comprehension. [Urquhart and Weir \(2014\)](#) highlight goal-setting as an important aspect of reading ability, noting that modifying one’s reading behavior based on the reading purpose is important. In other words, the type of reading task will influence the skills and behavior necessary to complete the task. This idea is expanded in the Reading as Problem Solving Model (RESOLV; [Rouet et al., 2017](#)) wherein a reader constructs a representation of a text with respect to the reading purpose and task at hand. Readers moderate the speed of reading and the level of attention to the text depending on whether the reader is skimming for gist (faster pace, global attention), scanning for details (faster pace, local attention), reading for informational purposes (slower pace, global attention), or having processing difficulty (slower pace, local attention) ([Carver, 1997](#); [Grabe, 2009](#)). Understanding these factors and how this is elicited by reading tasks is important for designing effective measures of reading comprehension ([Alderson, 2000](#); [Borsboom, 2005](#)).

However, it is difficult to observe reading behavior, let alone strategic reading. Part of why the previous debate about how L2 reading and whether it was more derived from L2 proficiency or L1 literacy came from this methodological difficulty in observing reading behavior. Reading abilities of either order have been difficult to measure directly, and as such, cognitive validity of reading tests could only be indirectly examined. That is until more sophisticated methods for tapping into cognitive processes of reading, such as eye-tracking became available ([Conklin et al., 2018](#)). Now, behavior related to both lower-order and higher-order reading abilities can be somewhat more directly observed.

Eye-tracking in second language acquisition

Observing reading processes and their contribution to successful comprehension has been a goal of Second Language Acquisition (SLA) research, but historically there have been few means by which to observe cognition in real time. Investigations into the processes which lead to successful comprehension have been usually been *post-hoc* in nature, but concurrent methods, such as eye-tracking, have become more commonplace ([Godfroid, 2019](#)). The utility of eye-tracking methods in investigating SLA rests on the assumption of the Eye-Mind Hypothesis ([Just and Carpenter, 1980](#)) stating that “eye movements are over orienting responses that signal the alignment of attention with the object at the point of gaze” ([Godfroid, 2019](#), p. 23). Visual attention can give us insight into how readers allocate cognitive resources to text. Although eye-tracking in reading is often restricted to processes related to local word parsing, there has also been attention paid to how Eye-tracking data can inform us about high-order reading cognition. For example [Yeari et al. \(2017\)](#) utilized eye-tracking methods to find that readers pay more or less attention to peripheral information depending on their reading purpose. [Dirix et al. \(2020\)](#) found that having readers

engage with a text for informational purposes elicited shorter overall reading times and shorter fixations than when readers engaged with a text for studying purposes, and that these differences were increased for L2 text-reading. They additionally found that students could compensate for slower processing with more overall attention to the text. Huang et al. (2022) examining Chinese L2 English learners' reading of texts with unfamiliar words. They found that working memory and duration of first fixation affected how readers processed unfamiliar words. Comprehension performance was affected by the longer duration of first fixation on unfamiliar words, yet unfamiliar word fixation affected comprehension less for learners who demonstrated higher working memory capability. This result demonstrates that successful reading can involve compensation for one weakness in reading with another resource.

Less attention has been paid to real-time reading behavior during L2 reading comprehension assessment. Bax and Chan (2019) measured second language English readers' eye-movements during reading test completion, finding that more successful readers made shorter fixations on average and paid more attention to areas of text based on relevance. In studies by Prichard and Atkins (2016, 2019), L2 English readers were found to underutilize strategic reading when they had time pressure to complete a reading task. Readers who were able to consciously apply strategic reading to their task did better in their comprehension. Outside of these studies, little research has been conducted on L2 reading assessment, especially with the analysis of interactions between components of reading ability in mind, but it is clear that eye-tracking can provide an avenue to understanding reading behaviors in relation to comprehension ability for L2 learners (Conklin et al., 2018).

Research questions

The goal of this study was to investigate whether differences in real-time reading behavior, as measured using eye-tracking, uniquely impacts second language reading comprehension performance, and to investigate interactions between reading behavior and other individual differences. Specifically,

- (1) To what extent do online reading behaviors predict variance in reading comprehension scores beyond that predicted by offline measures of individual cognitive and noncognitive differences (logic, memory, motivation, proficiency)?
- (2) To what extent do linear models reveal compensatory effects within individual differences impacting comprehension outcomes?

Methods

The data for this study involved second language English readers completing three sequential reading comprehension tasks each while reading one of a pool of six texts. During reading task completion, an eye-tracker recorded reader behavior. Each of the aspects of data collection and analysis are described below.

Participants

The data for this study was collected from 102 international students (graduate and undergraduate, with ages ranging between 19 and 52) at a large university in the southeastern United States as part of a larger study on second language reading assessment. The students represented a wide range of language backgrounds, including Mandarin, Spanish, Korean, Telugu, Cantonese, Urdu, Vietnamese, and 21 other language groups. Participants had spent an average of 4.67 years in an English-speaking environment, with an average of 5.1 years of English classroom experience.

Texts

The reading procedure involved reading three texts from a pool of six texts. The six texts were all passages from high school science textbooks on the following topics: "biotechnology and DNA," "the compound microscope," "chemical properties of water," "the science of hunger," "the psychology of making choices," and "attitudes and roles." Texts ranged from 315 to 350 words, and each consisted of four paragraphs. The texts were selected based on their similarity in terms of lexical and syntactic complexity, as well as their intended reading level of US high school grade 10 (Flesch Kincaid reading level is reported in [Supplementary Appendix A](#)). Although there is an inherent advantage in comprehension for any examinees with background knowledge on each particular topic, the texts were selected from introductory writings on the topic and reviewed by a panel of three applied linguists for broad approachability.

Tasks

Three reading comprehension tasks were completed by participants during the eye-tracking procedure. Each task reflected an oft-used second language reading test format along the spectrum of selected-response to constructed-response. The tasks were a multiple-choice (MC) reading task (selected-response, discrete-point scoring), a cloze task (constructed response, discrete-point scoring), and a summary task (constructed response, human scored). The MC task for each text involved answering five questions: one main-idea question, two detail questions, and two inferencing questions. Each question had three answer choices. Questions were presented to the right of the text and participants could see the text and questions at the same time without scrolling or leaving the screen.

The cloze task involved reading the text, but with 15 words replaced by blanks. There was no word bank to fill in the blanks, and participants needed to use comprehension processes to reconstruct the text. Words were blanked using a rational deletion method (Kleijn, 2018) targeting a content word or coherence-maintaining word every 15 words rather than a random or systematic deletion method to ensure that the task focused on comprehension processes as much as possible. Cloze tasks were scored by human raters so that near synonyms could be accepted

as correct answers. Scoring was otherwise objectively rated based on an answer key.

The summary task asked readers to produce a 100-word summary, or a “brief account” (Seidlhofer, 1990), of the text for a hypothetical fellow student who did not read the text. The provision of a specific audience and task encouraged summarizers to focus on content transmission and not linguistic copying and recall. As with the MC task, the task pane in which examinees typed their summaries was presented to the right of the text so the examinees could navigate between text and task without scrolling or changing screens. Summaries were scored by human raters for level of detail, evidence of mental modeling, and adherence to the task. Each text is presented in [Supplementary Appendix A](#).

Eye-tracking metrics

Readers’ real-time reading behavior was recorded with an ASL EyeTrac 6 device. Participants were seated two feet from a computer screen as they completed the reading comprehension tasks, keeping their head in a stable position using a chin rest. Each participant was calibrated with a practice exercise to ensure accuracy of fixations to within 0.2 inches before recording began. Fixation location and duration data were gathered by the eye-tracking device, along with length of *saccades* (jumps between fixations). Fixations were considered to be any pause in eye-movement >100 ms (Manor and Gordon, 2003). Lines of text and paragraphs were designated using *post-hoc* areas of interest (AOIs). Further AOIs were marked for each task area.

Various metrics were derived from the raw data which are relevant for understanding text-level reading behavior. The derived metrics are “late” processing measures, which reflect integrating of larger portions of text. These contrast with “early” processing measures, primarily focused on individual words and phrases. The metrics calculated in this study are average saccade length, total numbers of text fixations per word in reading text area and task areas, average fixation durations on the text and in task areas, and average fixations per word per dwell in AOIs. Unique for the assessment context, the number of transitions between a fixation on text and a fixation on a task area was calculated. Metrics related to rereading were also gathered, but they were largely multicollinear with total fixations per word, indicating that text level reading in an assessment setting naturally involves a great deal of rereading. Eye-tracking metrics were further evaluated for normality and text topic effects. These analyses are not reported in detail and were merely performed to ensure the assumptions were met for subsequent analyses. The metrics utilized in analyses are presented in [Table 1](#).

Although there was a time limit for the overall data collection procedure of 90 min, there was substantial variance in the amount of time taken to complete the individual reading tasks, so for each task, the eye-tracking metrics were checked for multicollinearity with reading time. The following metrics were found to be multicollinear ($r \geq 0.7$) with reading time and were excluded from further analysis: transitions in the cloze task ($r = 0.739$), text fixations per word in the cloze task ($r = 0.896$), task fixations per word in the cloze task ($r = 0.729$), text fixations per word in the MC task ($r = 0.762$), and task area fixations per word in the MC task (r

$= 0.744$). No fixation metrics were multicollinear with reading time for the summary task.

Individual differences

Considering the large number of cognitive factors which impact comprehension aside from eye-movement behavior, data from individual differences were gathered to understand what moderating effects might occur on how attention impacts task performance in reading assessment.

Language proficiency

Academic reading ability in a second language depends heavily on general grammatical knowledge and vocabulary size. Due to the diverse background of the participants, no standardized measure of proficiency could be gathered *a priori* for all participants, so an 18-item gap-fill c-test was developed to target morpho-syntax and academic vocabulary. The test involved 18 sentences with a word which was left half blank. The test is based on the productive orthographic vocabulary size tests (Laufer and Nation, 1999) which have been found to strongly predict reading comprehension in a second language (Cheng and Matthews, 2018).

Reading speed

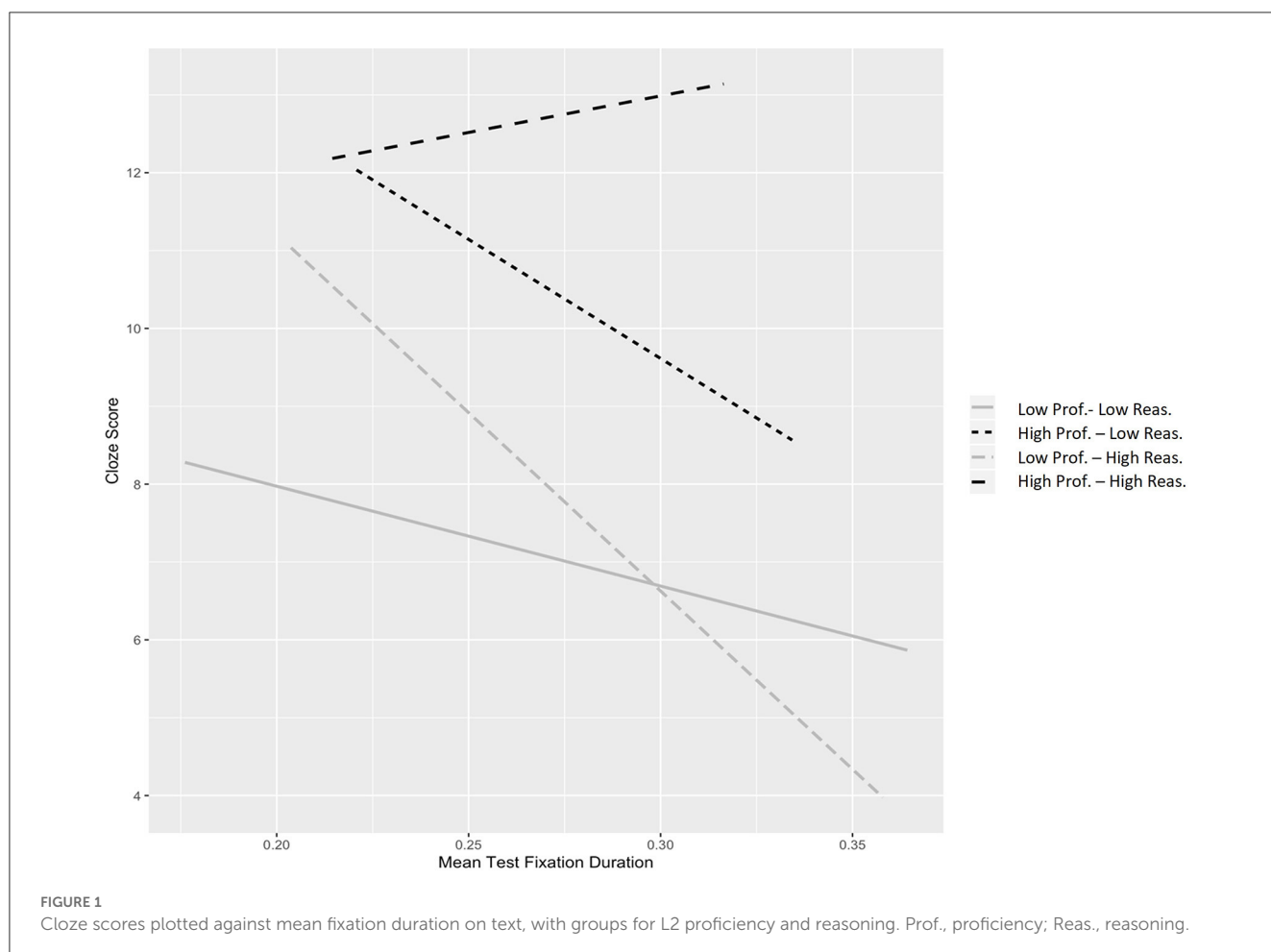
Reading fluency is an important lower-order literacy skill (Grabe, 2009; Gauvin and Hulstijn, 2010; Stoller et al., 2013), which has been found to be connected to reading behavior in monolingual data (Taylor and Perfetti, 2016). Reading fluency was here operationalized as reading speed in words per minute during a silent reading of a 12th grade-level academic text with 375 words about geology. The participants were asked comprehension questions afterward to ensure the participants read intentionally but the questions were not scored.

Reading motivation

Motivation is an important factor in understanding academic reading comprehension (Wigfield and Guthrie, 1997; Schaffner and Schiefele, 2013). A survey was developed to measure reading motivation and was administered before the reading trials. All items were discrete-point, using a 5-point Likert scale, and included 10 items. Five items measured intrinsic motivation to read, and five items measured extrinsic motivation to read. Intrinsic motivations include personal reasons such as enjoyment or personal enrichment, and extrinsic motivations include practical reasons such as career-usefulness of reading texts or social engagement through reading. These items were derived from previous surveys of motivation (Wigfield and Guthrie, 1997; Ryan and Deci, 2000). A confirmatory factor analysis was used to investigate the two-factor nature of the survey, resulting in a significant model ($\chi^2 = 58.23$, $p = 0.006$). However, only the intrinsic motivation questions reliably factored together in a unified construct, and so the intrinsic motivation metric was featured subsequent modeling of comprehension. The entire motivation survey is presented in [Supplementary Appendix C](#).

TABLE 1 Description and operationalization of eye-tracking measures.

Measure	Purpose for measurement	Target area	Operationalization notes
Fixations on text per word	Global, careful reading	Entire text area	Average of all fixations made on the reading text in a given trial
Mean length of saccade	Global reading	Entire trial area	Average absolute distance between sequential fixation coordinates throughout a trial
Mean fixations per line dwell	Linear, local reading	Line areas of interest	Average count of fixations per dwell across dwells in line AOIs. Controlled for number of words in AOI
Mean fixations per paragraph dwell	Local, careful reading	Paragraph areas of interest	Average count of fixations per dwell across dwells in paragraph AOIs. Controlled for number of words in AOI
Mean duration of fixations on text	Careful reading	Entire text area	Average time (ms) of fixations in any text area of interest. Controlled for size of AOI
Mean duration of fixations on task	Careful reading, Task integration	Task areas of interest	Average time (ms) of fixations in any task area of interest. Controlled for size of AOI
Fixations on task per word	Task integration	Task areas of interest	Average of all fixations made on the task areas in a given trial. Size of the areas in the respective tasks is controlled for
Number of gaze transitions between text and task	Task integration, global reading	Text and task areas of interest	Raw count of saccades which moved from a text area of interest to a task area of interest



Reasoning

Logical reasoning, or inductive reasoning, has been predictive of reading comprehension ability in previous research (Klauer and Phye, 2008). This facet of reasoning specifically refers to the ability

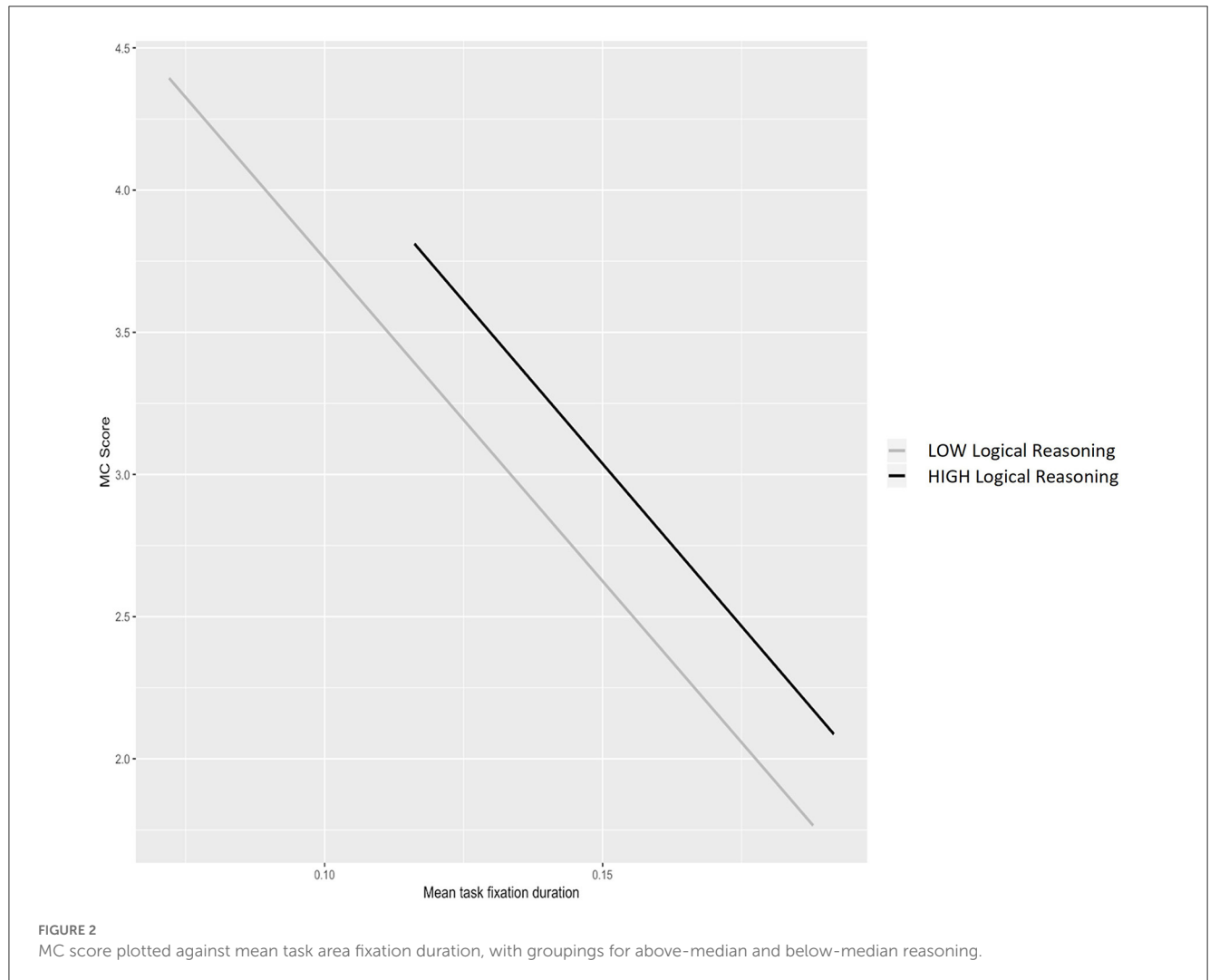
to extrapolate information from patterns. For this study, inductive reasoning was measured using a 10-item incomplete series test where participants saw a pattern of three shapes and selected the best of four options to complete the sequence.

TABLE 2 Linear regression model to predict cloze task scores.

Predictor	B	SE	t	p-value	r ²	Δr ²
Intercept	-0.021	0.072	-0.293	0.770		
L2 proficiency × reasoning × mean text fixation duration	0.151	0.070	2.150	0.034*	0.021	
L2 proficiency	0.663	0.074	8.959	<0.001*	0.442	0.421
Reasoning	0.278	0.078	3.551	0.001*	0.511	0.069
Mean text fixation duration	-0.222	0.072	-3.099	0.003*	0.559	0.048

B, standardized coefficients.

*Significant at p < 0.05.



Working memory

Working memory has been found to contribute to reading comprehension in monolingual readers (Cain et al., 2001; Calvo, 2005; Carretti et al., 2009) and multilingual readers (Alptekin and Erçetin, 2010; Lipka and Siegel, 2012; Erçetin and Alptekin, 2013; Joh and Plakans, 2017). Working memory was measured using a 2-back test, where participants were shown a series of simple images. Participants compared the image on screen to the image which they saw two images previously, deciding if they were the same within 1 s. They saw a total of 35 images, among which 15 2-back matches

were randomly distributed in the sequence of pictures. Participants were scored by the percent of correct responses.

Scoring

Each participant's responses were scored in a task-appropriate manner. MC task responses were scored automatically by key, and a score of 0 to 5 was assigned to each test-taker. Trained raters scored the cloze tests with an answer key using an acceptable

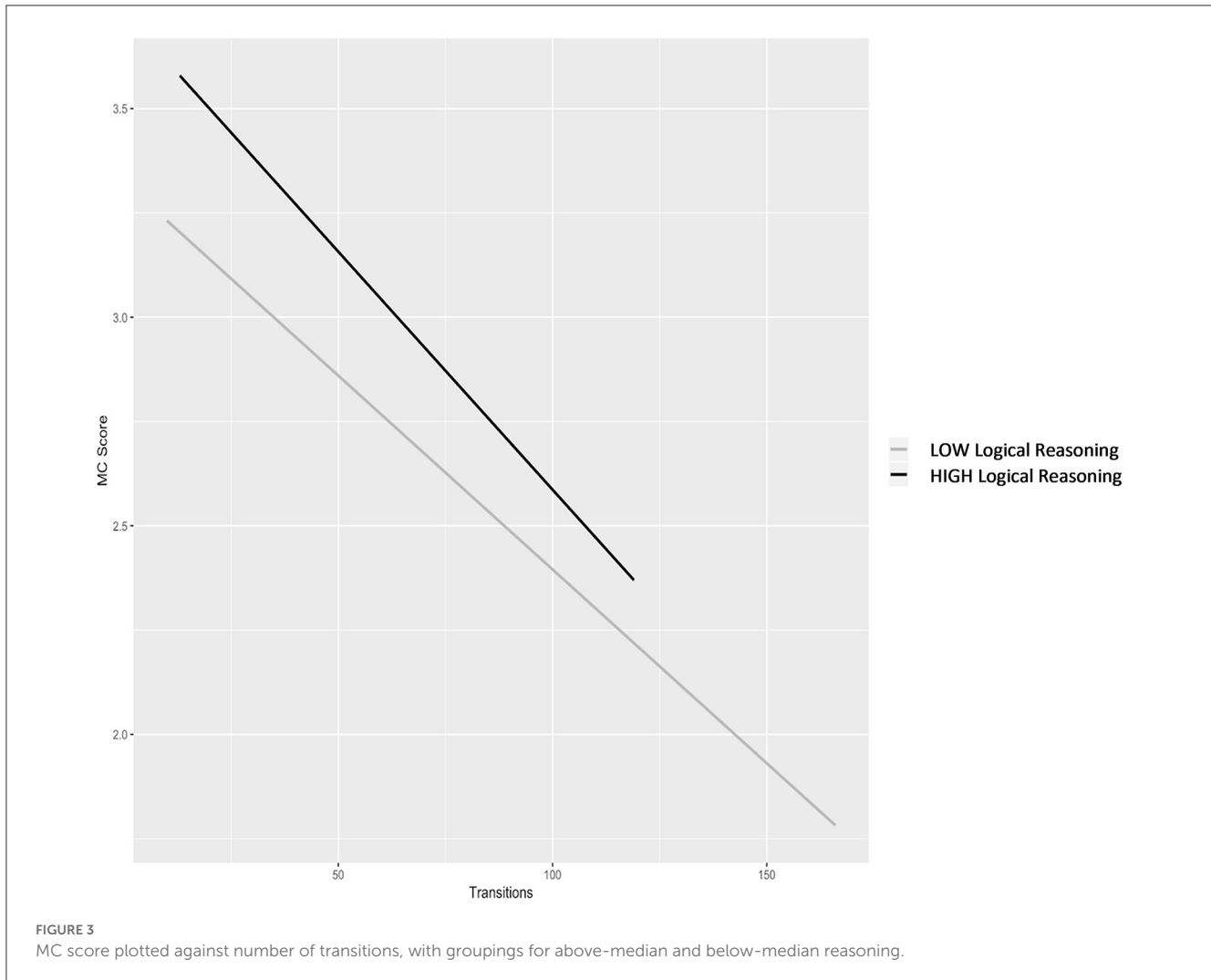


TABLE 3 Linear regression model to predict MC task scores.

Predictor	B	SE	t	p-value	r ²	Δr ²
Intercept	<0.001	0.092	0.000	1.00		
Mean fixation duration (task)	-0.347	0.092	-3.766	<0.001*	0.135	
Transitions	-0.280	0.092	-3.036	0.003*	0.213	0.078

B, standardized coefficients.
*Significant at p < 0.05.

response scoring method. Each cloze blank had an intended response based on the source text, but scorers also accepted near-synonyms. Each correct response to a blank in the passage was given a point, for a score range of 0 to 15 points. Trained raters also scored the summary tasks. The raters consisted of a pool of seven applied linguists. Summaries were rated using an analytic rubric developed by the researcher (see [Supplementary Appendix B](#) for the full summary rating guidelines). This rubric was developed based on constructs in [Taylor \(2013\)](#) used for rating summaries. The constructs include content accuracy, level of modeling (distinguishing between main ideas and subordinate details), task

completion, and language quality. Only accuracy, modeling, and task completion were considered as part of the comprehension score, with the language score being used to control for productive language ability and ensure raters did not factor linguistic aspects into their content scores. The language score component was only included on the rubric to mitigate the effect of raters' judgments of productive language quality in their assessment of reading comprehension.

Each summary was given a score out of 4 for each construct, and each summary was rated by at least two raters. If ratings from the two raters were misaligned in any category by more than one point, a third rater was called. Only 8.5% of ratings resulted in a third rater's adjudication, and no fourth ratings were necessary. The summary ratings were analyzed for reliability using Multifaceted Rasch Analysis ([Linacre and Wright, 2002](#); [Linacre, 2023](#)). Although the complete results of such an analytic measure are too voluminous to report here, importantly, the rubric constructs demonstrated independence with high separation reliability of 0.9, and the raters each exhibited acceptable fit, ranging from 0.72 to 1.12. This is within the acceptable range of model fit of 0.5 to 1.5 ([Linacre, 2023](#)), indicating good internal consistency among the raters.

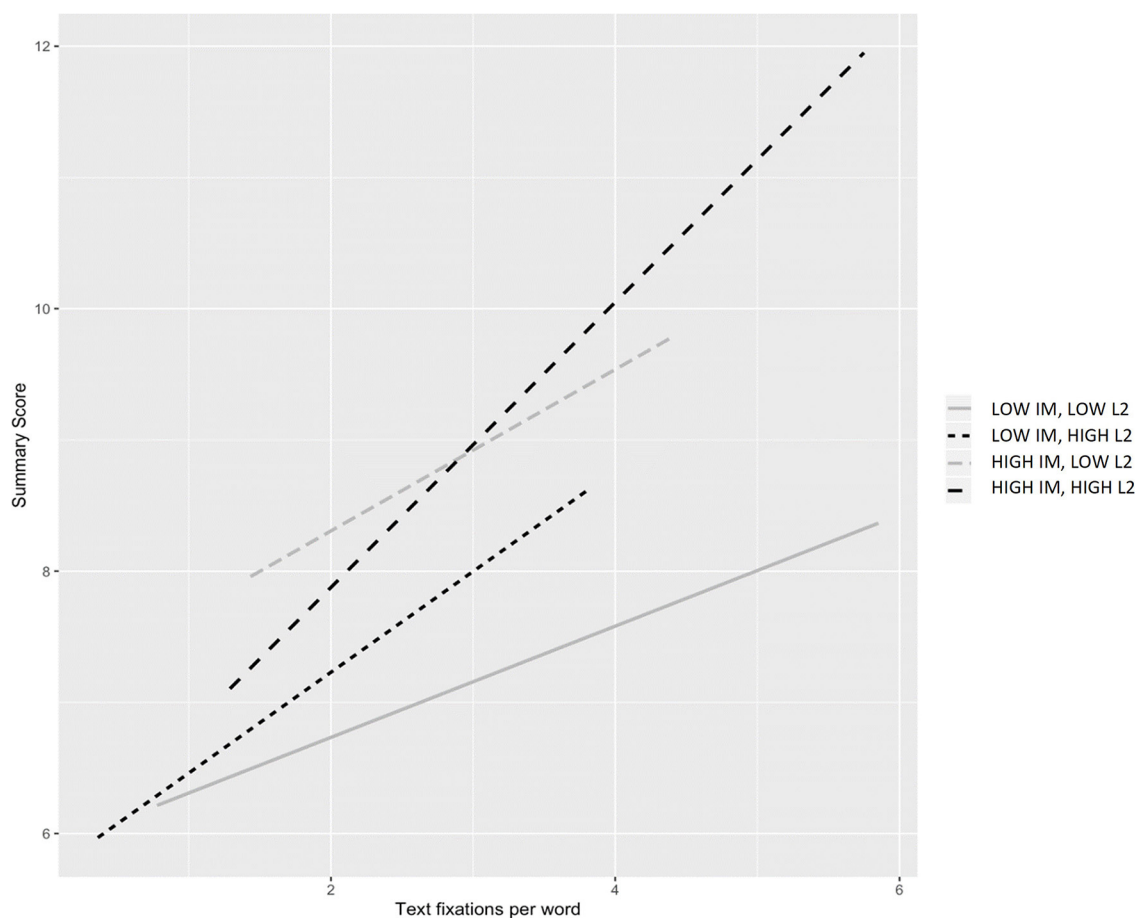


FIGURE 4

Summary score plotted against text fixations per word, with groupings for above-median and below-median motivation and L2 proficiency. IM, intrinsic motivation; prof., L2 proficiency.

The average of the closest two was used as the final score for each construct, and an additional Total Comprehension score was calculated as the sum of the accuracy, modeling, and task completion ratings. This total score was the score used as the dependent variable in summary modeling analyses.

Analyses

Three linear models were constructed to predict comprehension score in each task, using predictors of eye-tracking metrics along with individual differences which exhibited meaningful correlation with scores. A separate linear model was developed for each reading task. Correlations were calculated between each pair of metrics and with task scores. Eye-tracking and individual differences metrics which had significant and at least a weak correlation with score, were included in a linear regression model to predict score.

Results

This section will cover the results of the analyses described in the previous section on eye-movement and reading

comprehension. Comprehension scores for each of the different reading tasks were predicted with unique models, the construction of which began with examination of correlations. Based on correlations, eye-tracking metrics with at least a weak significant correlation with scores were selected for linear regression modeling. Similarly, individual difference metrics at least weakly significantly correlated with score were included as well. Text topic was included as a control variable.

Predicting cloze scores

One eye-tracking metric was found to correlate with cloze scores: mean fixation duration on text ($r = -0.306$). The correlation was negative, implying faster eye-movement via lower fixation durations was related to higher performance. Two individual differences were found to significantly correlate with cloze scores: L2 proficiency ($r = 0.630$) and logical reasoning ($r = 0.212$). The metrics were not correlated with each other or with average fixation duration on text.

Before constructing the linear model, visual inspection of the three variables was conducted to ascertain the presence of interactions. Figure 1 shows cloze scores along the y-axis, with

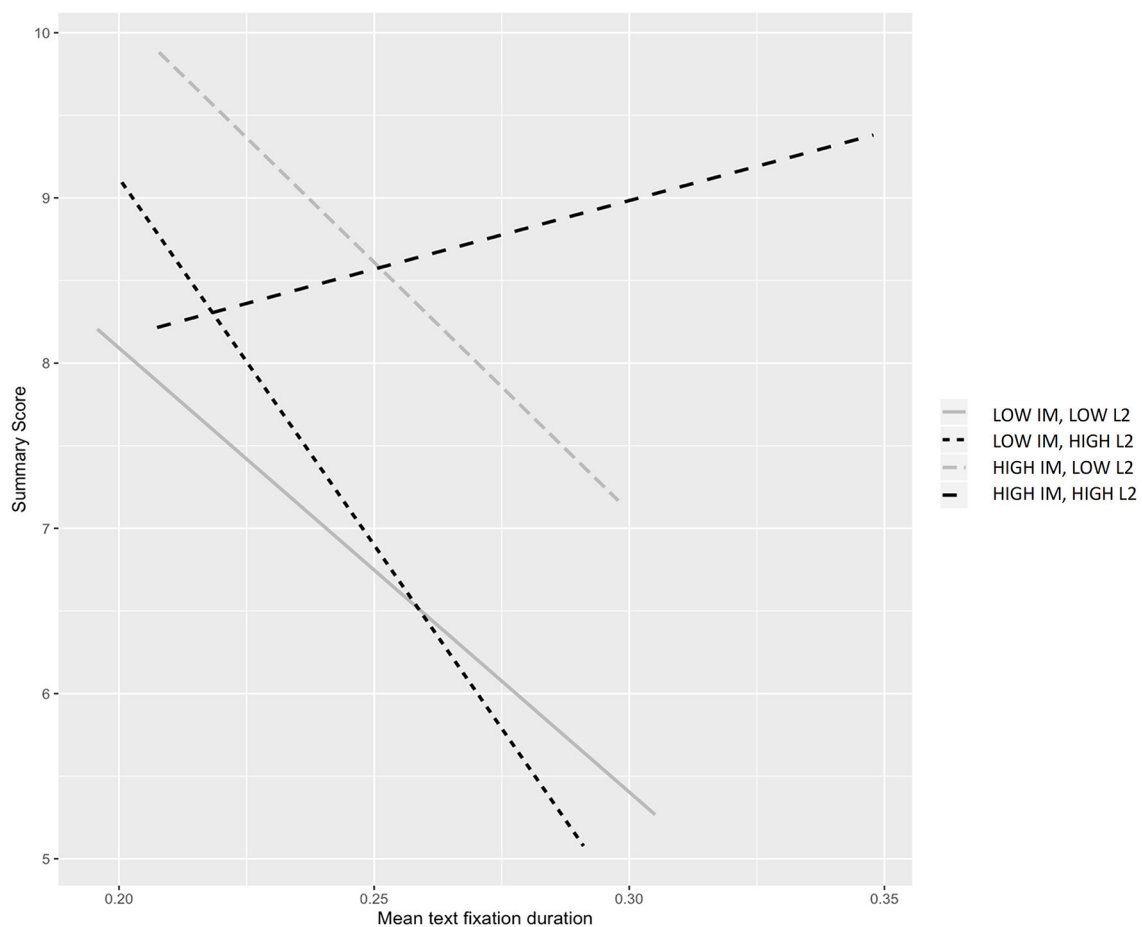


FIGURE 5

Summary score plotted against mean text fixation duration, with groupings for above-median and below-median motivation and L2 proficiency. IM, intrinsic motivation; prof., L2 proficiency.

mean fixation duration on text along the x-axis, and groupings for L2 proficiency level and reasoning level (each split into two groups around the median). The different slopes of the mean text fixation duration fit lines between proficiency levels and reasoning levels indicate a possible interaction effect. As such, these interactions were included in the linear modeling.

Variables were standardized for the linear model, and a linear model with three variables as well as on three-way interaction was constructed. The model was found to be significant, $F_{(4,94)} = 27.64$ ($p < 0.001$), and a description of the model is presented in Table 2. The model was found to have a large effect size, explaining 55.9% of variance in scores. Average fixation duration on text, as well as interactions with individual differences, was found to be uniquely account for variance in the model, though the effect size is very small. L2 proficiency and reasoning were positive predictors of score, and average fixation duration was a negative predictor, implying that shorter fixations related to higher scores. The interaction variable is more complex, but when interpreted alongside visual presentation of data in Figure 1, it can be seen that when both L2 proficiency and reasoning are above average, the negative impact of fixation duration reverses somewhat, i.e., readers ability to make fast fixations is less important when reasoning and L2 proficiency are high. This effect is small, but still indicates that these metrics may have a compensatory effect between them.

Predicting MC scores

Two eye-tracking metrics were found to correlate with MC scores: transitions between text and task areas ($r = -0.293$), and mean fixation duration on the question area ($r = -0.379$). Each of the correlations were negative, implying fewer transitions and shorter fixations on the question area were related to higher MC performance. Only a single individual difference metric was found to significantly correlate with MC scores, logical reasoning ($r = 0.221$). Reasoning was not significantly correlated with any eye-tracking metrics.

Before constructing the linear model, visual inspection of the two variables was conducted to ascertain the presence of interactions. Figure 2 through Figure 3 show MC scores along the y-axis, with eye-tracking metrics along the x-axis, and groupings for reasoning (split into two groups around the median). The participants were split into groups for above median or below median in reasoning to make the plots reader friendly, and this grouping is not used in further analysis. The similar slopes of the average fixation duration and transitions fit lines between reasoning levels indicates that higher reasoning scores trend with higher comprehension scores, and there is likely little to no interaction effect between the reasoning and eye-movement behavior.

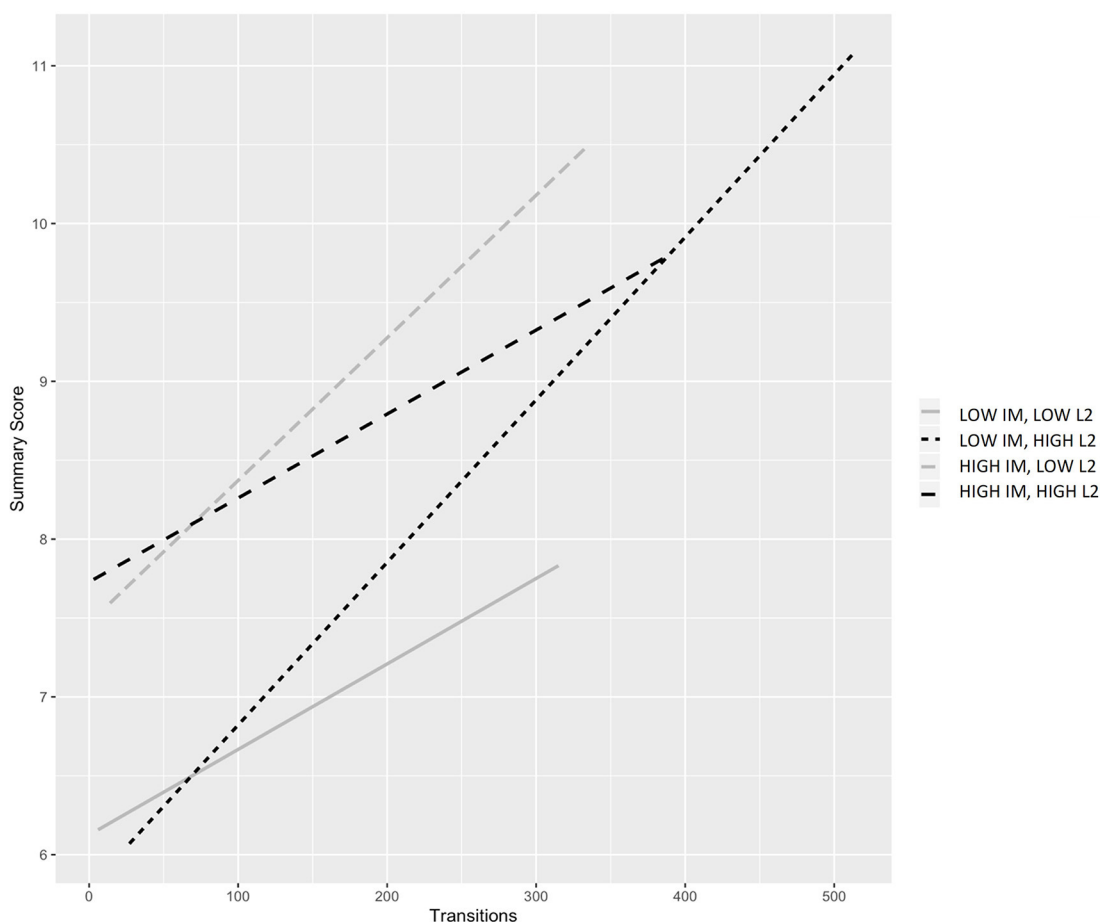


FIGURE 6

Summary score plotted against number of transitions, with groupings for above-median and below-median motivation and L2 proficiency. IM, intrinsic motivation; prof., L2 proficiency.

The linear regression model for MC score included as predictors reasoning, average duration of fixation on questions, and transitions. Of the included predictors, mean fixation duration on questions and transitions were significant predictors, but reasoning and any interaction variables were not. These effects were thus removed from the model. The final 2-predictor model was found to be significant, $F_{(2,96)} = 12.583$ ($p < 0.001$) and Table 3 contains a description of the model. The model had a moderate effect size in predicting score, with $r^2 = 0.213$. Mean fixation duration on questions was the most significant predictor, showing that making shorter fixations on the question area contributed to higher scores. Transitions was also a significant predictor, with fewer transitions being predictive of higher score.

Predicting summary scores

Three eye-tracking metrics were found to correlate with summary scores: transitions between text and task areas, this time positively correlated ($r = 0.302$), fixations per word in text area ($r = 0.364$), and mean fixation duration on the text ($r = -0.214$). As in the cloze data, mean duration of fixations on the text was

negatively correlated with summary score. Fixations per word on text was significantly correlated with transitions ($r = 0.477$), but there were no multicollinear variables. Two individual difference metrics were found to significantly correlate with summary scores: L2 Proficiency ($r = 0.297$) and Intrinsic Motivation ($r = 0.345$).

Before constructing the linear model, visual inspection of the three variables was conducted to ascertain the presence of interactions. Figures 4–6 show Summary scores along the y-axis, with eye-tracking metrics along the x-axis, and groupings for individual differences split around the median. The participants were split into groups for above median or below median in reasoning to make the plots reader friendly, and this grouping is not used in further analysis. Each graph reveals interaction effects between the eye-tracking metrics and the individual differences, but the most can be found in the graph for mean fixation duration. Here, mean fixation duration normally has a negative correlation with summary score, yet at higher levels of both motivation and L2 proficiency, the relationship between fixation duration and summary score is positive. These interactions are further explored for significance in the linear regression model.

The linear regression model for Summary score included as predictors intrinsic motivation, L2 proficiency, fixations per word

TABLE 4 Linear regression model to predict summary task scores.

Predictor	B	SE	t	p-value	r ²	Δr ²
Intercept	0.004	0.083	0.043	0.966		
Intrinsic Mot. × L2 proficiency × mean fix. duration	0.253	0.087	2.902	0.005*	0.058	
L2Proficiency × mean fix. duration	0.211	0.078	2.707	0.008*	0.063	0.005
Intrinsic motivation	0.188	0.086	2.188	0.031*	0.149	0.086
L2 proficiency	0.177	0.087	2.044	0.044*	0.207	0.058
Fixation per word (Text)	0.385	0.086	4.498	<0.001*	0.331	0.124
Mean fixation duration (text)	-0.275	0.089	-3.098	0.003*	0.397	0.066

B, standardized coefficients.

*Significant at $p < 0.05$.

on text, mean fixation duration on text, and number of transitions. Number of transitions and the interactions with it were not found to be significant to the model and were removed. The final model was found to be significant, $F_{(6,92)} = 9.641$ ($p < 0.001$). Table 4 contains a description of the model. The effect size of the model was large, with about 39.7% of the variance explained for summary scores ($r^2 = 0.397$). The three-way interaction with L2 proficiency, motivation, and mean fixation duration was found to be significant and a positive predictor of summary scores, where mean fixation duration alone was a significant negative predictor. The stronger of the two predictors was mean fixation duration alone, indicating that the positive interaction does not mean readers with stronger proficiency and motivation necessarily benefit from longer fixations, but rather mitigate slower fixations with their other abilities. A positive pairwise interaction between L2 proficiency was also significant in the model, but not to the extent of the three-way interaction. This still further shows the strength of L2 proficiency to compensate for more rapid fixations.

In addition to mean fixation duration, three other main effects were found to be significant. Fixations per word on text was the most meaningful predictor, indicating higher numbers of fixations predicted higher summary scores with a moderate effect size. High motivation was a moderate positive predictor as well, and L2 proficiency had a main effect, but it was not as impactful on score as its interaction effects with text duration.

Discussion

The online reading behavior measured in this study was used to understand its impact on reading comprehension and interactions with individual differences across various reading assessment tasks. Each reading task elicited a different linear model to predict comprehension scores using individual differences and eye-tracking metrics. These are briefly summarized below.

Score on the cloze was related to L2 proficiency, reasoning, and efficiency of fixations. Shorter fixations on text areas was predictive of cloze score, with a small but meaningful effect size ($\Delta r^2 = 0.048$), though this was not as meaningful as the predictive effects of L2 Proficiency (to a large extent) and reasoning. The three way interaction between these variables indicated that at higher

levels of proficiency and reasoning, the effect of fixation efficiency diminished as other skills could compensate.

The model predicting score on the MC task was much weaker, with two eye-movement measures related to processing the question area of the text being meaningful in the model. Having shorter fixation durations on the questions and fewer transitions between question and text predicted higher comprehension scores. Though there was a possible interaction between reasoning and number of fixations, with higher reasoning scores relating to lower fixations, neither these main effects nor this interaction was significant in the score model.

The model predicting summary task scores included multiple predictors, with motivation, proficiency, and fixations positively predicting summary scores with at least a weak effect size, and mean fixation duration negatively predicted scores. There was again an interaction, with longer fixation durations no longer having a negative impact on score at higher levels of proficiency and/or motivation. Readers with higher motivation appear to be able to compensate for the impact of slower processing ability on comprehension with more L2 linguistic resources.

To answer the first research question, *to what extent do online reading behaviors predict variance in reading comprehension scores beyond that predicted by other individual differences*, we can look at the appearance of eye-tracking main effects in the models of comprehension for each reading task. For each reading task model, a fixation duration metric was found to predict scores, with shorter average fixations predicting higher score. This is in line with previous research which showed that skilled readers make short, efficient fixations (Ashby et al., 2005; Bax, 2013; Krieger et al., 2016). The summary task was distinct from the cloze and MC tasks in that an eye-tracking metric positively predicted scores. For the summary task, a greater number of fixations on the reading text was predictive of higher summary scores with a medium effect size. It is possible that the summary task pushes readers to build a more detailed mental model of the text and is more cognitively demanding, so more fixations are necessary. This is attested in Bax (2013) who found eye-movement behavior related to higher-order processing in summary comprehension tasks.

In relation to the second research question, *to what extent do linear models reveal compensatory effects within individual differences impacting comprehension outcomes*, interactions were present in two models of reading comprehension. The results from this study align with previous research which asserts that

readers can compensate for certain weaknesses in reading ability by utilizing other related skills (Stanovich, 1980; McNeil, 2012). McNeil's (2012) framework made predictions about how readers at different levels would rely on strategic, literate, or linguistic resources. Although the current study did not seek to ascertain which aspect of skills would impact comprehension most at different levels of reading, we nonetheless established that L2 language ability, reading behavior, and strategic abilities have unique contribution to reading comprehension, and readers can compensate for weaknesses in one skill with strengths in another. The specific compensations related to reading efficiency, where efficiency was less critical for comprehension when readers had higher L2 proficiency and/or another skill (logical reasoning for the cloze task and motivation for the summary task). This deviates slightly from previous research which found interaction effects with eye-tracking metrics on reading comprehension. In Huang et al. (2022), working memory was found to be a significant predictor of comprehension, and was able to compensate for the effect of unfamiliar words which caused slower processing. However, the Huang et al. (2022) study was looking at smaller texts with shorter reading times, so the results of the current study extend our understanding of how measures of efficient processing materialize at different lengths of text. For longer texts and tasks allowing simultaneous access to text and task, working memory may not be the most predictive cognitive measure, and may not compensate for late-measure eye-tracking metrics as measured in this study.

Conclusion

This study has taken a novel look at how reading behavior, measured through eye-tracking, differed across reading tasks in terms of impact on task performance. Beyond furthering our understanding of the second-language reading process, there are implications for language teaching and testing as well. It is worth acknowledging as teachers that readers benefit from learning various aspects to reading, from refining language proficiency to practicing extensive reading for speed to engaging in reasoning and motivation-enhancing tasks. Since there is variance in how different abilities contribute to comprehension performance across tasks, it is also worth teaching developing readers goal-setting strategies to help them compensate for the demands set by their reading purpose. For example, reading for discrete information as in the cloze and MC tasks demands quick, efficient reading, but reading for global comprehension as in the summary task required more comprehensive attention to the text. Being able to moderate one's approach to reading in different tasks is critical.

These findings must be taken in light of the study's limitations. Previous research (Cook and Wei, 2019) has advised against drawing direct connections between eye-tracking metrics and underlying processes. This is especially true for the current study which utilized very coarse-graining eye-tracking metrics. Fixations per word and average fixation duration are both general measurements based on participants' entire trial of reading data. More attention to areas of interest and phrasal/word-level eye-tracking information could provide more to the picture of eye-movement behavior's contribution to comprehension. Further research is needed to better understand how finer shades of measurement of fixation duration impacts comprehension and

relates to other individual differences. It is also necessary to state that while we observed the impact of reading efficiency in this study, we were not able to ascertain whether readers consciously engaged in faster or slower reading as part of an active reading strategy. More research is needed to connect eye-movement behaviors to conscious engagement in specific types of reading strategy activation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Georgia State University Institutional Review Board [University Research Services & Administration (URSA)]. The patients/participants provided their written informed consent to participate in this study.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Data collection for this research would not have been possible without support from the Georgia State University Adult Literacy Research Center Dissertation Support Grant.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2023.1176986/full#supplementary-material>

References

- Alderson, J. C. (2000). *Assessing Reading (Atlanta Library North 4 LB1050.46.A43 2000)*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511732935
- Alptekin, C., and Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *J. Res. Read.* 33, 206–219. doi: 10.1111/j.1467-9817.2009.01412.x
- Ashby, J., Rayner, K., and Clifton, C. (2005). Eye movements of highly skilled and average readers: differential effects of frequency and predictability. *Q. J. Exp. Psychol.* 58, 1065–1086. doi: 10.1080/02724980443000476
- Bachman, L. F., and Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: OUP Oxford.
- Bax, S. (2013). The cognitive processing of candidates during reading tests: evidence from eye-tracking. *Lang. Test.* 30, 441–465. doi: 10.1177/0265532212473244
- Bax, S., and Chan, S. (2019). Using eye-tracking research to investigate language test validity and design. *System* 83, 64–78. doi: 10.1016/j.system.2019.01.007
- Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annu. Rev. Appl. Linguist.* 25, 133–150. doi: 10.1017/S0267190505000073
- Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511490026
- Cain, K., Oakhill, J. V., Barnes, M. A., and Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Mem. Cognit.* 29, 850–859. doi: 10.3758/BF03196414
- Calvo, M. G. (2005). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learn. Individ. Differ.* 15, 53–65. doi: 10.1016/j.lindif.2004.07.002
- Carretti, B., Borella, E., Cornoldi, C., and De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: a meta-analysis. *Learn. Individ. Differ.* 19, 246–251. doi: 10.1016/j.lindif.2008.10.002
- Carver, R. P. (1997). Reading for one second, one minute, or one year from the perspective of rauding theory. *Sci. Stud. Read.* 1, 3. doi: 10.1207/s1532799xssr0101_2
- Cheng, J., and Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Lang. Test.* 35, 3–25. doi: 10.1177/0265532216676851
- Conklin, K., Pellicer-Sanchez, A., and Carroll, G. (2018). *Eye-tracking: A Guide for Applied Linguistics Research*. Cambridge: Cambridge University Press. doi: 10.1017/9781108233279
- Cook, A. E., and Wei, W. (2019). What can eye movements tell us about higher level comprehension? *Vision* 3, 45. doi: 10.3390/vision3030045
- Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., and Duyck, W. (2020). Reading text when studying in a second language: an eye-tracking study. *Read. Res. Q.* 55, 371–397. doi: 10.1002/rrq.277
- Erçetin, G., and Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: implications for second-language reading comprehension. *Appl. Psycholinguist.* 34, 727–753. doi: 10.1017/S0142716411000932
- Gauvin, H. S., and Hulstijn, J. H. (2010). Exploring a new technique for comparing bilinguals' L1 and L2 reading speed. *Read. Foreign Lang.* 22, 84–103.
- Godfroid, A. (2019). *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*. London: Routledge. doi: 10.4324/9781315775616
- Grabe, W. (2009). *Reading in a Second Language: Moving from Theory to Practice*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139150484
- Huang, L., Ouyang, J., and Jiang, J. (2022). The relationship of word processing with L2 reading comprehension and working memory: insights from eye-tracking. *Learn. Individ. Differ.* 95, 102143. doi: 10.1016/j.lindif.2022.102143
- Joh, J., and Plakans, L. (2017). Working memory in L2 reading comprehension: the influence of prior knowledge. *System* 70, 107–120. doi: 10.1016/j.system.2017.07.007
- Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354. doi: 10.1037/0033-295X.87.4.329
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge University Press.
- Klauer, K. J., and Phye, G. D. (2008). Inductive reasoning: a training approach. *Rev. Educ. Res.* 78, 85–123. doi: 10.3102/0034654307313402
- Kleijn, S. (2018). *Clozing in on Readability. How Linguistic Features Affect and Predict Text Comprehension and On-line Processing*. Dutch: LOT, Netherlands Graduate School.
- Koda, K. (1988). Cognitive process in second language reading: transfer of L1 reading skills and strategies. *Second Lang. Res.* 4, 133–156. doi: 10.1177/026765838800400203
- Koda, K. (1990). The use of L1 reading strategies in L2 reading: effects of L1 orthographic structures on L2 phonological recoding strategies. *Stud. Second Lang. Acquis.* 12, 393–410. doi: 10.1017/S0272263100009499
- Krieger, M., Bartl-Pokorny, K. D., Pokorny, F. B., Einspieler, C., Langmann, A., Körner, C., et al. (2016). The relation between reading skills and eye movement patterns in adolescent readers: evidence from a regular orthography. *PLoS ONE* 11, e0145934. doi: 10.1371/journal.pone.0145934
- Lauffer, B., and Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Lang. Test.* 16, 36–55. doi: 10.1191/026553299672614616
- Linacre, J. M. (2023). *Facets Computer Program for Many-facet Rasch Measurement*, version 3.85.1. Beaverton, OR: Winsteps.Com.
- Linacre, J. M., and Wright, B. D. (2002). Construction of measures from many-facet data. *J. Appl. Meas.* 3, 486–512.
- Lipka, O., and Siegel, L. (2012). The development of reading comprehension skills in children learning English as a second language. *Read. Writ.* 25, 1873–1898. doi: 10.1007/s11145-011-9309-8
- Manor, B. R., and Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *J. Neurosci. Methods* 128, 85–93. doi: 10.1016/S0165-0270(03)00151-1
- McNeil, L. (2011). Investigating the contributions of background knowledge and reading comprehension strategies to L2 reading comprehension: an exploratory study. *Read. Writ.* 24, 883–902. doi: 10.1007/s11145-010-9230-6
- McNeil, L. (2012). Extending the compensatory model of second language reading. *System* 40, 64–76. doi: 10.1016/j.system.2012.01.011
- Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730
- Prichard, C., and Atkins, A. (2016). Evaluating L2 readers' previewing strategies using eye tracking. *Read. Matrix* 16, 110.
- Prichard, C., and Atkins, A. (2019). Selective attention of L2 learners in task-based reading online. *Read. Foreign Lang.* 31, 269–290.
- Rouet, J.-F., Britt, M. A., and Durik, A. M. (2017). RESOLV: readers' representation of reading contexts and tasks. *Educ. Psychol.* 52, 200–215. doi: 10.1080/00461520.2017.1329015
- Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi: 10.1006/ceps.1999.1020
- Schaffner, E., and Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: does text accessibility during comprehension testing make a difference? *Learn. Individ. Differ.* 26(Supplement C), 42–54. doi: 10.1016/j.lindif.2013.04.003
- Seidlhofer, B. (1990). Summary judgments: perspectives on reading and writing. *Read. Foreign Lang.* 6, 413–424.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Read. Res. Q.* 16, 32–71. doi: 10.2307/747348
- Stoller, F. L., Anderson, N. J., Grabe, W., and Komiyama, R. (2013). Instructional enhancements to improve students' reading abilities. *English Teach. Forum* 51, 2–11.
- Taylor, J. N., and Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Read. Writ.* 29, 1069–1103. doi: 10.1007/s11145-015-9616-6
- Taylor, L. (2013). *Testing Reading through Summary: Investigating Summary Completion Tasks for Assessing Reading Comprehension Ability*. Cambridge: Cambridge University Press.
- Urquhart, A. H., and Weir, C. J. (2014). *Reading in a Second Language: Process, Product and Practice*. London: Routledge. doi: 10.4324/9781315841373
- Wigfield, A., and Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *J. Educ. Psychol.* 89, 420–432. doi: 10.1037/0022-0663.89.3.420
- Yeari, M., Elentok, S., and Schiff, R. (2017). Online and offline inferential and textual processing of poor comprehenders: evidence from a probing method. *J. Exp. Child Psychol.* 155(Supplement C), 12–31. doi: 10.1016/j.jecp.2016.10.011