languages

Systematic Review

A Systematic Review of Eye-Tracking Technology in Second Language Research

Xin Hu and Vahid Aryadoust *

National Institute of Education, Nanyang Technological University, Singapore 637616, Singapore; nie21.hx3882@e.ntu.edu.sg

* Correspondence: vahid.aryadoust@nie.edu.sg

Abstract: Eye-tracking has become increasingly popular in second language (L2) research. In this study, we systematically reviewed 111 eye-tracking studies published in 17 L2 journals to explore the application and replicability of eye-tracking technology in L2 research. The results revealed eight areas of application of eye-tracking in L2 research, among which grammar and vocabulary were the most frequently examined lines of inquiry. We also identified three types of cognitive mechanisms investigated in L2 eye-tracking studies: attention, higher cognitive processes, and cognitive load. Attention was predominantly measured via fixation temporal indices, while higher cognitive processes were frequently measured by using fixation count and fixation temporal measures. In addition, the measures adopted to assess cognitive load mainly depended on the task type. Finally, with respect to the replicability of the studies, transparent reporting practices were evaluated based on 33 features of replicable studies. We found that more than 95% of the reviewed studies reported less than 70% of the information essential for future replication studies. We suggest that the reporting of the information critical to conducting replicable L2 eye-tracking research needs improvement in transparency and completeness. The implications of this study are discussed.

Keywords: eye-tracking; replicability; reporting practices; second language research; systematic review

check for updates

Citation: Hu, Xin, and Vahid Aryadoust. 2024. A Systematic Review of Eye-Tracking Technology in Second Language Research. *Languages* 9: 141. https://doi.org/10.3390/languages9040141

Academic Editor: Jon Clenton

Received: 21 December 2023
Revised: 6 March 2024
Accepted: 7 March 2024
Published: 12 April 2024

1. Introduction

Eye-tracking is a real-time data collection method that monitors participants' eyes while they read texts or view images displayed on a computer screen (Aryadoust and Ang 2021; Godfroid 2019). Over the past two decades, an increasing number of second language (L2) studies have adopted eye-tracking to explore how learners process L2s in the context of various language tasks (Conklin and Pellicer-Sánchez 2016; Godfroid 2019; Roberts and Siyanova-Chanturia 2013). This tool has been frequently used to capture the underlying mechanisms of L2 processing, such as word recognition (e.g., Marian and Spivey 2003) and grammatical processing (e.g., Cunnings et al. 2017; Roberts et al. 2008), and the acquisition of vocabulary and grammar (e.g., Godfroid et al. 2013; Issa and Morgan-Short 2019; Lee and Révész 2020; Pellicer-Sánchez et al. 2021a). Among L2 skills, reading and listening are the two main areas of investigation in the eye-tracking literature (e.g., Aryadoust 2020; Aryadoust and Foo 2023; Aryadoust et al. 2022; Bax 2013; Kho et al. 2023; Kim and Grüter 2021; Low and Aryadoust 2023; Mitsugi 2017; Traxler et al. 2021; Zhou et al. 2020).

There have been two recent research syntheses (Abdel Latif 2019; Godfroid 2019) exploring important topics of L2 eye-tracking research, both showing that eye-tracking has been applied to multiple L2 domains. Godfroid (2019) also provided a list of eye-tracking measures used in L2 research so far. To further this synthesizing work, in the present study, we conducted a systematic review to determine the types of eye-tracking measures used across research domains. More specifically, the first objective of our study is to identify research areas that L2 scholars have focused on, along with measures of gaze behaviors adopted in each research area. Our study differs from the review of Godfroid (2019) in that we developed a framework to classify eye-tracking measures, informed by eye-tracking research across research fields (e.g., Lai et al. 2013). In particular, with the advancement of knowledge and technology, there are emerging eye-tracking measures that might be useful but not yet being applied in L2 research. The study will thus provide insights into how eye-tracking has been used; enable researchers to be aware of the gaps in each research strand; and importantly may open new research frontiers in L2 eye-tracking studies.

Another gap in understanding is the types of cognitive mechanisms explored by L2 researchers and the corresponding eye-tracking measures collected for analysis. Gaze behaviors, which directly indicate visual attention, can reflect a variety of higher cognitive processes, including comprehension and learning (Alemdag and Cagiltay 2018; Rayner 2009; Son et al. 2021). As discussed below, understanding the utility of eye-tracking measures in representing L2 cognitive processes would be of particular interest and use for future research. To this end, this study aims to synthesize a research-based cognitive framework of gaze dynamics, thereby highlighting the utility of eye-tracking measures for diverse research applications. It will also show how eye-tracking measures have been and may be interpreted in various ways across research designs, thus indicating that caution should be exercised in drawing inferences of cognition based on eye-tracking measures.

Finally, informed by ongoing calls for the replicability of primary research, the third objective of this study is to critically evaluate the current state of replicability of L2 eye-tracking studies with a particular focus on reporting practices. Without transparent reporting and sufficient information on empirical research, the replicability of eye-tracking studies would be threatened. With this in mind, improving the methodological rigor and transparent reporting practices by establishing more robust reporting standards is called for in this study. In particular, with the growing interest in methodological quality in L2 research (Norris et al. 2015), it is important to promote higher methodological standards for the use of eye-tracking in L2 research.

2. Eye-Tracking and Cognitive Mechanisms

Gaze behaviors, from the eye fixation to the pupil size, can be modulated by cognition, suggesting that cognitive processes can influence the movement and direction of the eyes during the processing of visual information (Holmqvist et al. 2011; Rayner et al. 2006). Applied eye-tracking researchers have leveraged this notion to investigate various cognitive operations. For example, visual attention, normally conceptualized as a selective process, can be measured by using eye-tracking technology (Aryadoust and Ang 2021; Son et al. 2021). It is generally believed that the eyes are tightly linked to attention when performing cognitively demanding tasks, such as reading (Rayner 2009). Therefore, studies in L2 and applied linguistics have explored, for instance, the allocation of attention to different linguistic constructions while reading (e.g., Godfroid et al. 2013; Issa and Morgan-Short 2019) and attention distribution in the processing of multimodal materials (e.g., Aryadoust and Foo 2023; Low and Aryadoust 2023; Pellicer-Sánchez et al. 2020).

Higher cognitive processes (Wang et al. 2006), such as comprehension and learning, can also be inferred from gaze behaviors (Lai et al. 2013; Rahal and Fiedler 2019; Rayner et al. 2006). Most eye-tracking researchers in applied linguistics and beyond assume that there is a close relationship between what is fixated on and what is processed in the mind, a concept referred to as the "eye-mind assumption" (Conklin et al. 2018; Godfroid 2019; Just and Carpenter 1980; Rayner 2009). Thus, eye movements can be indicative of the moment-to-moment cognitive processes that occur during language comprehension and/or learning. Taking reading as an example, word frequency has been shown to influence word recognition and elicit variations in eye-movement patterns: fixation durations on high-frequency words are shorter than on low-frequency words (Rayner et al. 2006). In addition to eye movements, recent research has also shown that pupil diameters and blinks are valuable indicators of cognitive processing (Eckstein et al. 2017).

However, it is important to acknowledge the limits of eye-tracking, too. Gaze measures cannot be directly linked to specific higher-level cognitive processes and there are bounds as to how much insight eye-tracking can provide (Conklin et al. 2018). Eye-tracking measures are in fact "simply measures of visual behavior (e.g., gaze position and related movements)" (King et al. 2019, p. 6). Therefore, eye-tracking data alone does not show which cognitive processes are operated in the brain (Holmqvist et al. 2011). Put another way, it is not possible to use the gaze behavior recording itself to determine the precise moment at which a word is recognized or integrated to form a coherent understanding of the text (Conklin et al. 2018). Thus, researchers should determine what theoretical variables are operationalized by the collected eye-tracking metrics and properly link gaze behaviors to their assumed underlying cognitive processes.

3. Eye-Tracking Measures

Eye-tracking systems can provide the location, sequence, and duration of eye movements in the areas of interest (AOIs) as well as information on the changes in the pupil and the blinks of the eyes in real time (Holmqvist et al. 2011). There are multiple ways to conceptualize eye-tracking measures. Lai et al. (2013) summarized eye-movement measures based on the scale of measurement used (temporal, spatial, and count) and the type of eye movement (fixation, saccade, and mixed). Temporal measures quantify gaze behaviors temporally and are believed to provide insights into the points at which and for how long cognitive processing is undertaken, as well as into the processing load (Godfroid 2019; Lai et al. 2013). The spatial scale represents eye movements in a spatial dimension and is therefore concerned with locations, distances, directions, or sequences. Spatial measures can provide information as to where and how cognitive processing is undertaken. The count scale quantifies eye movements in terms of number, proportion, rate and/or frequency. Similarly, in SLA and bilingualism research, Godfroid (2019) grouped eye-tracking measures into three overarching categories: fixation, regression, and eye movement dynamics. According to Godfroid (2019), there are four subtypes of fixation-based measures, as follows: (1) fixation counts, probabilities, and proportions, (2) fixation duration, (3) fixation latency (e.g., the duration of time a participant takes before fixating on a specific area of interest), and (4) fixation location. In addition to fixations and saccadic eye movements, measures such as pupil size and blink rate have also emerged in recent studies to explore visual information processing and examine cognitive mechanisms (Eckstein et al. 2017). Drawing on the aforementioned frameworks, the current study groups the commonly used eye-tracking measures identified from language research and related fields of study into eight main categories: fixation, saccade, dwell (visit), regression, skip, pupillometry, blink, and gaze patterns (Table 1). The first four types are classified into the abovementioned three scales (Appendix A presents examples and definitions of the commonly used eye-tracking measures).

Table 1. Definitions of the Eye-tracking Measures.

Eye-Tracking Measure	Definition
Fixation	The periods of time during which the eyes remain stationary on a region.
Saccade	The rapid movements of the eyes between two consecutive fixations.
Dwell (visit)	The period of time during which a participant's gaze first enters an AOI until exiting that region.
Regression	The backward eye movements during reading.
Skip	The AOI that is never looked at by the participant.
Pupillometry	The fluctuations in the pupil's size and orientation.
Blink	The rapid closing and reopening of the eyelid.
Gaze patterns	The visualization of the temporal distribution and duration of eye movements.

4. Replicability and Research Reporting Practices

The replication of empirical studies constitutes a crucial method for verifying research findings, uncovering potential biases, and generalizing research findings to different conditions and populations (Makel and Plucker 2014). Although replication is believed to be essential for advancing scientific knowledge, there is a growing concern over the replicability of scientific research across research fields. Notably, the result of a *Nature*'s survey on 1567 researchers showed that over 70% of researchers tried and failed to replicate another scientist's study and that more than half of the researchers even failed to replicate their own experiments (Baker 2016). In response to such a "crisis", various attempts have been made to improve replicability, one of which relates to transparent reporting practices of scientific research.

Transparency in reporting practices involves providing sufficient information about the research procedure and key variables for others to understand, consume and replicate this study (Derrick 2016). Replication studies can be categorized into three types based on the change made to the methodology of the original study. Researchers can opt to duplicate the experimental procedure of the original study (direct replication), change specific facets of the methodology (partial replication) or adopt different methods (conceptual replication) (Makel and Plucker 2014; Marsden et al. 2018). Thus, transparent reporting of the original research is a prerequisite for successful replications.

In L2 research, multiple methodological syntheses have raised concerns regarding the incomplete reporting of instruments (Crowther et al. 2021; Derrick 2016), with there being "a history of inadequate reporting practices" (Marsden et al. 2018, p. 332). Similarly, there is a lack of transparency in reporting eye-tracking studies with respect to key aspects and features of eye-tracking research (e.g., the visual stimuli size, apparatus, eye-tracking data quality and algorithms) in many research fields, as found in the systematic reviews of eye-tracking studies in decision making research (Fiedler et al. 2020), communication science (King et al. 2019) and mathematics education research (Strohmaier et al. 2020). Specifically, Strohmaier et al. (2020) identified "large inconsistencies in the reporting of these methods" (p. 165), while Fiedler et al. (2020) found that many key elements of the eye-tracking research were omitted in reports regarding empirical studies in decision-making research. To promote transparent reporting practices of eye-tracking research, researchers have developed reporting guidelines for eye-tracking studies in several research fields (Carter and Luke 2020; Fiedler et al. 2020), but none of them are specialized for L2 research. This means that key variables of L2 research are neglected in these guidelines.

To our knowledge, there is no research synthesis that has examined the reporting practices of the currently available eye-tracking studies in L2 research. Despite the aforementioned concerns, there is a lack of empirical investigation into the replicability of L2 eye-tracking studies. It is on this basis that the present study aims to examine the extent to which elements critical to the carrying out of L2 eye-tracking studies have been reported in a transparent manner, and thus other researchers can replicate the original studies using the information provided.

5. The Present Study

Systematic reviews are gaining traction in the field of L2 research, because they have clear objectives and are methodologically sound, which are featured with systematicity, rigor, and transparency (Petticrew and Roberts 2008). The present study employs the systematic review approach to synthesize empirical eye-tracking studies in L2 research. This research synthesis sets out to explore where and how eye-tracking has been used in L2 research, including the areas of application, the cognitive mechanisms that have been investigated using eye-tracking, and the types of eye-tracking measures used in those areas and cognitive mechanisms. Moreover, it will critically evaluate the replicability of L2 eye-tracking studies from the perspective of transparent reporting practices. To guide the review, the research questions (RQs) of this study are as follows:

RQ1: What are the main areas of investigation in the L2 eye-tracking literature? What eye-tracking measures have been used in each research area?

RQ2: What types of cognitive mechanisms have been inferred from the eye-tracking measures collected?

RQ3: How replicable are the eye-tracking studies in L2 research?

6. Method

6.1. Study Identification

The dataset for this study was constructed through a sequential process following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Liberati et al. 2009; Moher et al. 2009) in December 2021. In line with the procedures used in previous research syntheses of L2 studies, this study focused on articles published in top-tier peer-reviewed journals because these journals have been considered as the primary means for disseminating high-quality L2 research (Marsden et al. 2018) (see Appendix B for the list of journals). To retrieve research papers that used the eye-tracking method in the selected journals, the Scopus database was chosen to conduct an electronic document search using the "source title" method because it is the largest available database of published studies (Aryadoust et al. 2021; Schotten et al. 2017). The search terms "eye tracking" and "eye movement" were applied using the "OR" Boolean operator. No limitation was set on the year of publications, but the language was restricted to English (Appendix C provides the Scopus search code). The initial search returned 154 articles.

All articles retrieved were screened with the inclusion and exclusion criteria (Table 2). A total of 111 journal articles were identified for coding and analysis (see Appendix D for the publications included in the systematic review). The PRISMA flow diagram (Moher et al. 2009) in Figure 1 documented the study selection process.

Table 2. Inclusion and Exclusion Criteria.

Inclusion Criteria: The Paper . . .	**Exclusion Criteria: The Paper** . . .
1. was published in the selected peer-reviewed journals;	1. was a book chapter, conference proceeding, or dissertation;
2. collected data from L2 (L3, L4, foreign language, artificial language) learners, educators, or materials;	2. did not include data from L2 learners, educators, or materials;
3. used the eye-tracking method;	3. did not use the eye-tracking method;
4. was a primary study that contained empirical data;	4. was secondary research, review, or commentary;
5. was published in English.	5. was inaccessible.

The final dataset consisted of 111 articles, published between January 2003 and December 2021. This dataset was substantially larger than those used in previous reviews of eye-tracking studies in L2 research (Abdel Latif 2019; Godfroid 2019), and could be considered as representative of the domain of interest. As shown in Figure 2, there has been a general upward trend in the number of papers using eye-tracking in L2 research, peaking at 25 in 2021. Since the publication year was not restricted in selection, it indicated that L2 studies that used the eye-tracking method emerged in the selected journals since 2003. This suggests that eye-tracking has become an established research method in L2 research and has gained more attention in recent years.

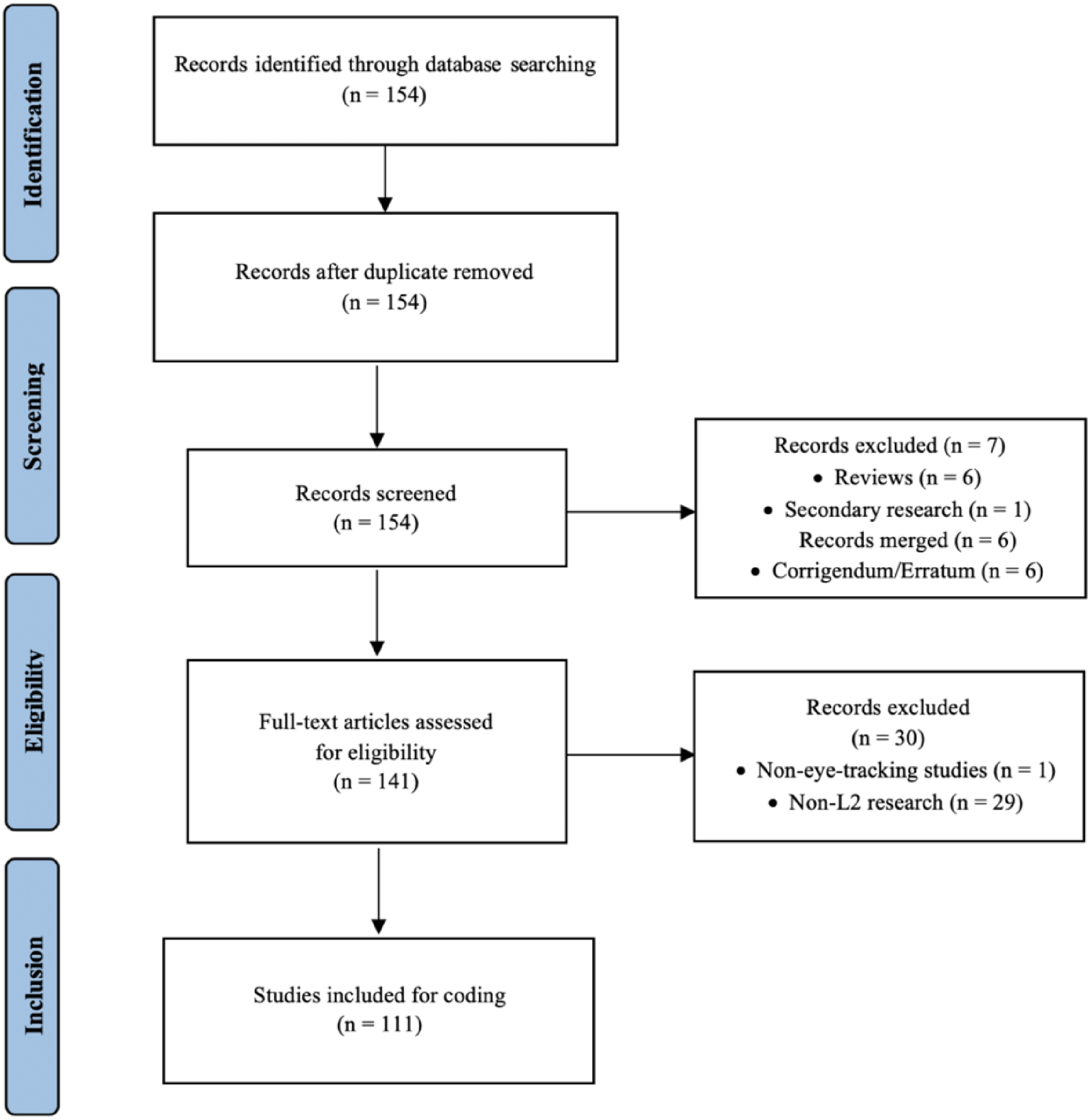


Figure 1. PRISMA Flow Diagram of Selection Strategy.

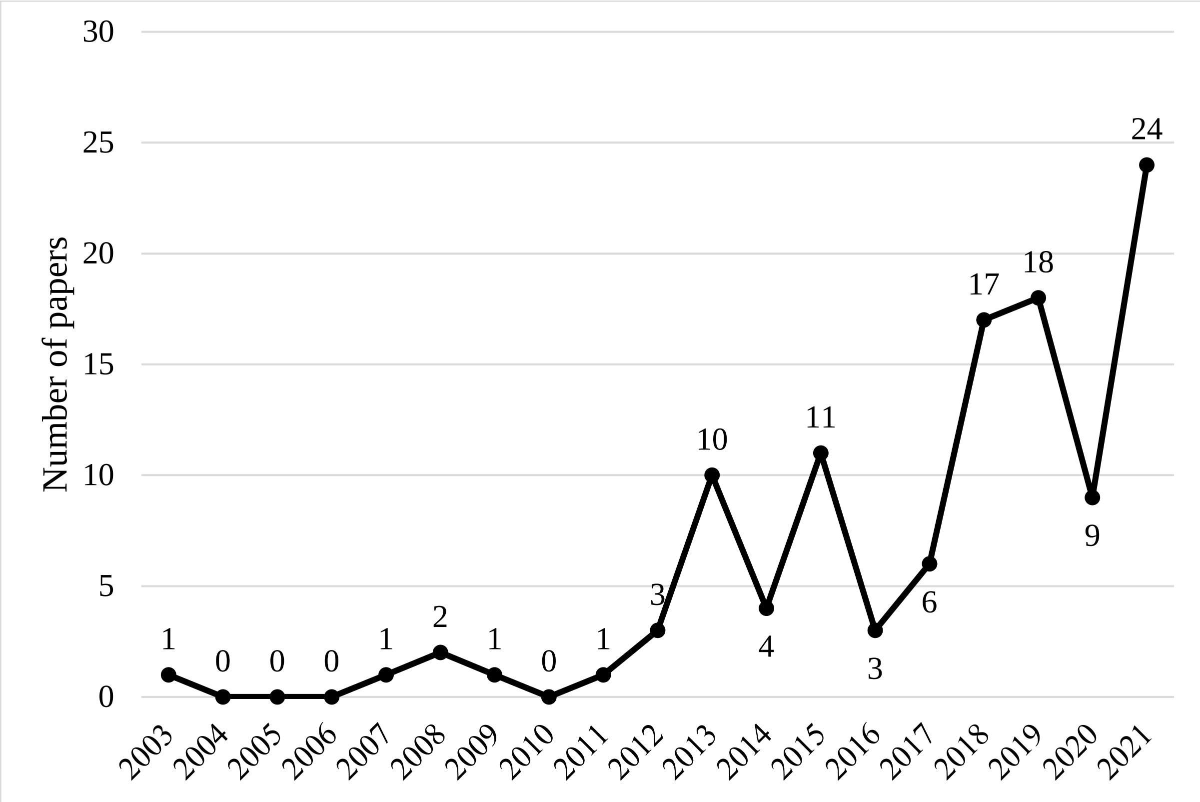


Figure 2. Line Graph Representing the Number of Papers in the Dataset over the Years.

6.2. Coding

A coding scheme (Appendix E) was developed to derive the information necessary from each study in the sample to address the research questions. In designing the coding scheme, the categories and variables were informed by three main resources: (a) research synthesis guidelines (e.g., Cooper et al. 2019), (b) previous research syntheses of L2 studies and eye-tracking studies (e.g., Crowther et al. 2021; Riazi et al. 2020; Strohmaier et al. 2020), and (c) the eye-tracking literature (e.g., Fiedler et al. 2020; Godfroid 2019; Holmqvist et al. 2011). The initial coding scheme was developed by the authors. Two independent reviewers, with expertise in eye-tracking, subsequently provided suggestions to revise and enhance the categories, variables, definitions, and values within the coding parameters. Thus, the final coding scheme was established through an iterative process in which the authors revised and piloted coding categories and variables, resolved disagreements and amended unclear codes and coding parameters.

Each study was subsequently coded for all the 45 items based on the coding scheme, and the 111 papers were coded by three coders. Firstly, the first author coded 30 papers, and another two independent reviewers each coded half of the coded papers (i.e., 15 papers each). Inter-coder reliability was examined for the coding of these 30 studies. Table 3 presents the inter-coder agreement rate for each variable in the coding scheme. The inter-coder agreement rate was 96.94%. Subsequently, the authors and reviewers discussed and resolved disagreements. Next, the first author coded 51 papers, and the two independent reviewers each coded 15 papers.

Table 3. Inter-Coder Agreement Rate for Each Variable.

The Variable	Inter-Coder Agreement Rate
Area of application	83.33%
Cognitive mechanism(s) inferred	80.00%
Eye-tracking measure	93.33%
Sample size	96.67%
Gender distribution	100.00%
Age	96.67%
L1	100.00%
Target L2	100.00%
L2 proficiency	100.00%
Neurological condition	93.33%
Visual condition	100.00%
Hearing condition	93.33%
Research site	93.33%
Visual sitimuli type	100.00%
Font type	100.00%
Font size	100.00%
Text spacing	100.00%
Image size	100.00%
Area of interests	100.00%
Commercial/non-commercial	100.00%
Type	86.67%
Brand/manufacturer	93.33%
Model	100.00%
Data sampling frequency	100.00%
Number of eyes tracked	90.00%
Head movement condition	100.00%
Display monitor	100.00%
Type of software used	96.67%
Name (and version)	96.67%
Eye data source	96.67%
Data quality	100.00%
Data interpolation	100.00%
Noise reduction	100.00%
Techniques for parsing eye movements	100.00%
Fixation threshold	100.00%
Velocity threshold	100.00%

6.3. Data Analysis

The analysis of the codes drew on descriptive statistics. The first research question was addressed by summarizing the research areas identified from the sample. After identifying the research areas, the frequencies and percentages of eye-tracking measure types used across research areas were calculated and summarized. Similarly, the second research question was answered by calculating the frequencies and percentages of eye-tracking measure types applied across the identified cognitive mechanisms. The third research question was intended to examine the reporting practices of the sample by analyzing the number of studies that reported each item. We also created a replicability index which was the function of A/B, where A = the amount of information provided and B = the amount of information needed for an exact replication. It should also be noted that in computing the denominator of the replicability index, we did not count in the sampling frequency and commercial/non-commercial type of the eye tracker, since these pieces of information can be found on the website of the products.

7. Results

7.1. Research Question 1: Research Areas and Eye-Tracking Measures Applied

The first research question was posed to illustrate the eye-tracking measures used across research areas. This was followed by an in-depth analysis of the distribution of eye-tracking research and the prevalence of eye-tracking measures used in each research subfield.

7.2. Span of L2 Eye-Tracking Studies

Figure 3 demonstrates the distribution of empirical L2 eye-tracking studies by research area. The categorization of studies was based on the primary area of research interest expressed by the researchers in the title, abstract, research aims, and research questions addressed by eye-tracking. Overall, eight main research areas and one mixed area emerged from the sample. Grammar (*n* = 27; 24.3%) is the most researched component, followed by vocabulary (*n* = 26; 23.4%), and reading (*n* = 17; 15.3%), while considerably less research attention has been directed to speaking (*n* = 2; 1.8%) and phonology (*n* = 2; 1.8%). In addition, as demonstrated in Table 4, several research areas consisted of two or more categories, indicating the span of the knowledge base in these research areas. For example, grammar studies tend to focus on grammar acquisition and instruction (e.g., Indrarathne and Kormos 2017; Wong and Ito 2018) and grammatical processing (e.g., Fujita and Cunnings 2021; Keating 2009), while listening research mainly focuses on predicting language processing (e.g., Kim and Grüter 2021; Mitsugi 2017), listening tests (e.g., Aryadoust et al. 2022; Suvorov 2015), and prosody (e.g., Connell et al. 2018; Wiener et al. 2021).

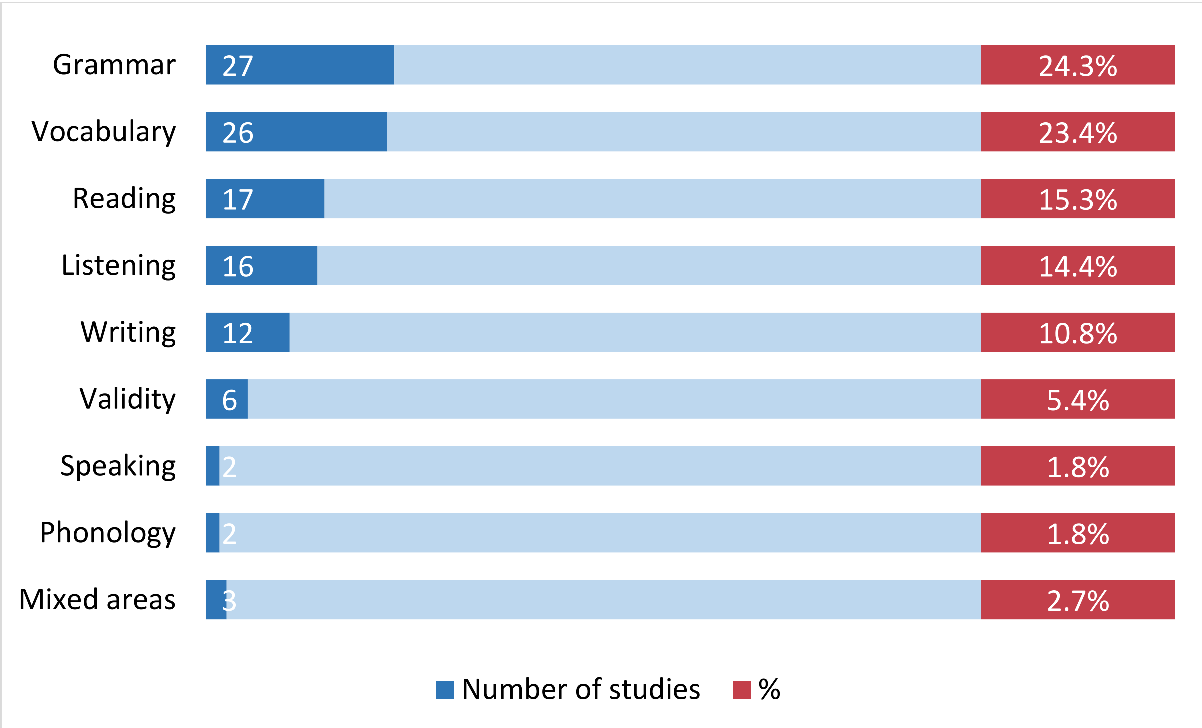


Figure 3. Distribution by the Research Area of Empirical Eye-Tracking Studies.

Table 4. Topics Investigated in L2 Eye-tracking Studies.

Research Area	Topic
1. Grammar	Grammar acquisition and instruction, grammatical processing
2. Vocabulary	Vocabulary acquisition and instruction, bilingual word recognition, formulaic language processing, conceptual transfer, and strategy use
3. Reading	Reading behavior, multimodal reading, and reading test
4. Listening	Predictive language processing, listening test, and prosody
5. Writing	Composing process, computer-mediated communication, feedback, and writing assessment
6. Validity	The validity of eye-tracking method, construct validity, and task validity
7. Speaking	Speaking test and event description
8. Phonology	Visual sonority and phoneme learning
9. Mixed areas	

7.3. Eye-Tracking Measures Used across L2 Research Areas

Table 5 presents the breakdown of the eye-tracking measures used in the L2 eye-tracking studies reviewed. The most frequently used measure was fixation temporal (n = 71; 64.0%), followed by fixation count (n = 54; 48.6%), dwell temporal (n = 17; 15.3%), and dwell count (n = 11; 9.9%). By contrast, fixation spatial (n = 2; 1.8%) and saccade count (n = 2; 1.8%) were the least frequently applied eye-tracking measures in the sample, while pupil dilation and blink were not employed in any of the studies.

Table 5. Breakdown of Eye-tracking Measures and Number of Studies.

Eye-Tracking Measure		Number of Studies	%
Fixation			
	Temporal	71	64.0%
	Count	54	48.6%
	Spatial	2	1.8%
Saccade			
	Temporal	0	0.0%
	Count	2	1.8%
	Spatial	3	2.7%
Dwell			
	Temporal	17	15.3%
	Count	11	9.9%
	Spatial	0	0.0%
Regression			
	Temporal	8	7.2%
	Count	8	7.2%
	Spatial	3	2.7%
Skip		8	7.2%
Blink		0	0.0%
Pupil dilation		0	0.0%
Gaze pattern		6	5.4%
Others		6	5.4%

We further investigated the types of eye-tracking measures applied within each research area. Details of the eye-tracking measure types used across the eight research areas and/or subareas are illustrated in Appendix F. Overall, the majority of the research areas and/or the subcategories showed a tendency in the field to use fixation temporal measures.

Only a minority of research topics used other measure types more frequently than fixation temporal measures, such as those in bilingual word recognition (in vocabulary) and listening research. Although 12 measure types were found to be employed in the dataset, each research area only utilized a subset of them. The most comprehensive coverage of measures occurred in reading behavior, validity, and mixed areas, wherein seven of the identified measure types were applied.

7.4. Research Question 2: Cognitive Mechanisms

The second research question focused on how L2 researchers used eye-tracking and gaze behaviors to understand different cognitive mechanisms. As previously discussed, eye-tracking provides a means to examine cognition. This capacity was used to study three main types of cognitive mechanisms in the reviewed studies: (1) attention, (2) higher cognitive processes, and (3) cognitive load. We found that 94 studies employed the collected eye-tracking measures to make inferences concerning participants' cognitive mechanisms, while the other 17 studies used eye-tracking to measure constructs such as viewing time and fluency that were not directly linked to any specific cognitive mechanisms by the authors of the studies (Figure 4).

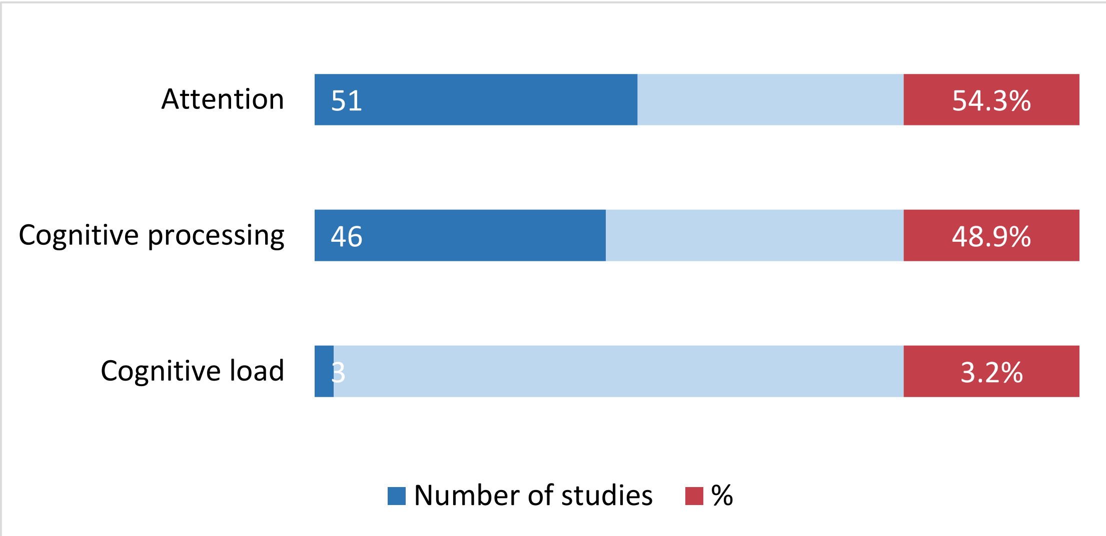


Figure 4. The Type of Cognitive Mechanisms Inferred from Eye-Tracking Measures (N = 94). *Note*: the percentage exceeds 100% as several studies exploring more than one type of cognitive mechanism were double-counted.

Attention emerged as the most studied cognitive mechanism in the sample. A total of 51 studies (54.3%) employed eye-tracking to measure, for example, the amount of attention, attentional distribution, and the allocation of attention. The measures used for analyzing attention spanned across eight eye-tracking measure types (Figure 5). The majority of these studies (n = 42; 82.4%) used fixation temporal measures, while 43.1% of the studies (n = 22) employed fixation count and 23.5% of the studies (n = 12) used dwell temporal measures. The other measure types included dwell count (n = 7; 13.7%), gaze pattern (n = 4; 7.8%), skip (n = 3; 5.9%), regression count (n = 3; 5.9%), and saccade spatial (n = 1; 2.0%), which were used less frequently.

The category of "higher cognitive processes" included 46 studies (48.9%), which used eye-tracking measures to infer the moment-to-moment cognitive processes underlying visual information processing, such as the language comprehension processes of lexical access, syntactic parsing, and predictive language processing. The reported measure types in the studies focused on higher cognitive processes spanned across 11 out of 12 measure types identified from the reviewed studies (Figure 6). Among these, fixation count measures (n = 27; 58.7%) were most frequently adopted, followed by fixation temporal (n = 22; 47.8%), regression temporal (*n* = 7; 15.2%), and regression count (*n* = 7; 15.2%) measures. The least applied measure types were fixation spatial (*n* = 1; 2.2%), saccade spatial (*n* = 1; 2.2%) and dwell count (*n* = 1; 2.2%) measures, each being reported in only one study. The types of eye-tracking measures used to infer cognitive load are presented in Appendix G.

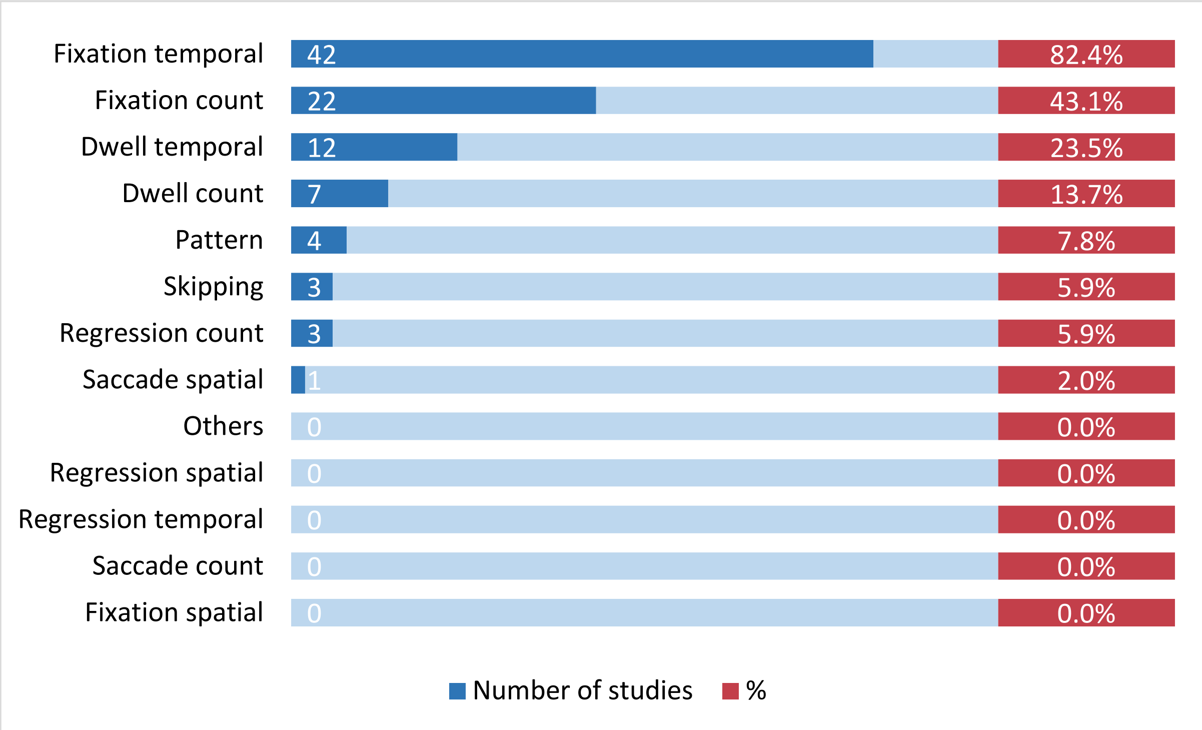


Figure 5. The Type of Eye-Tracking Measure Used to Infer Attention.

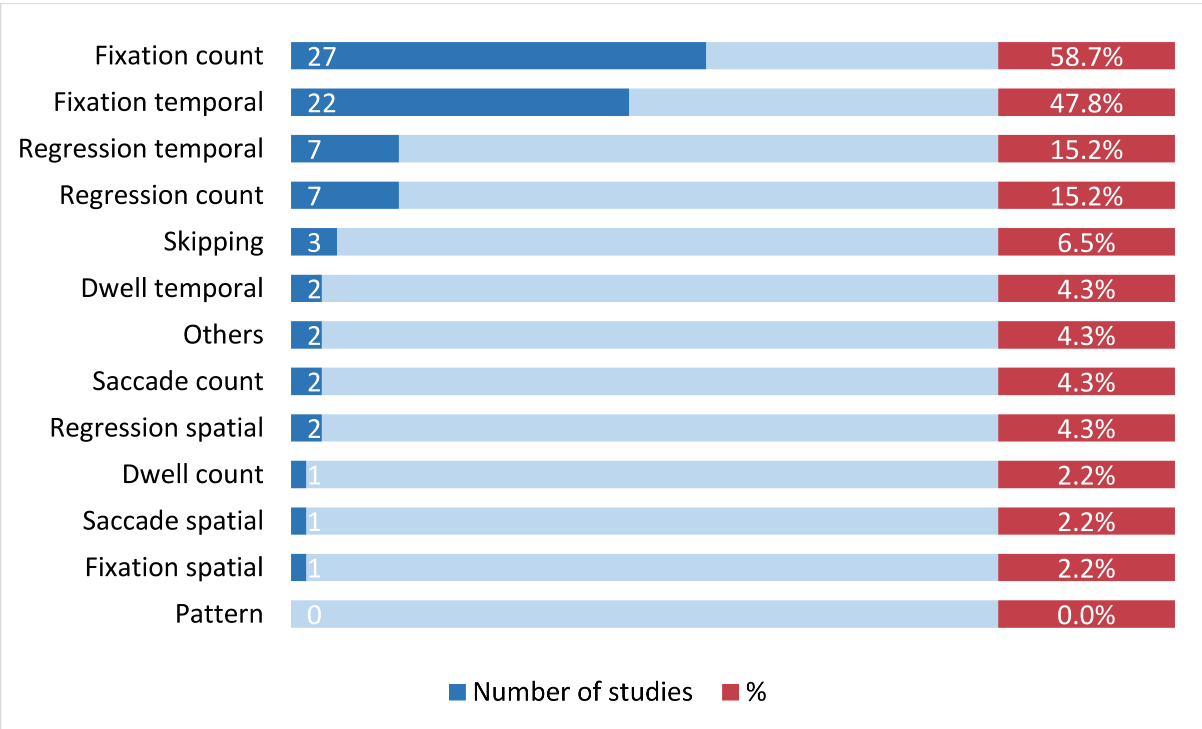


Figure 6. The Type of Eye-Tracking Measure Used to Infer Higher Cognitive Processes.

7.5. Research Question 3: Replicability of L2 Eye-Tracking Studies

The third research question of the study concerned the replicability of L2 eye-tracking studies, whereby we investigated the reporting practices of eye-tracking experiments in the sample. The dataset consisted of 121 separate L2 eye-tracking experiments documented across 111 publications, which were coded for 33 items in six categories associated with the study design, methodology, data collection, and eye-tracking data pre-processing procedures. Frequencies and percentages of different study contexts, participant demographics, apparatus, and data pre-processing procedures were calculated (see Appendices H and I). To provide a general overview of reporting completeness, Table 6 outlines the number of categories of information reported in the reviewed studies. It is apparent from this table that very few studies (*n* = 10; 8.3%) included all six categories of information (at least one item of each category) in the papers, indicating that over 90% of studies (*n* = 111; 91.7%) failed to specify at least one category of information. In addition, all reviewed studies reported more than two of the six categories, and nearly half of the studies (*n* = 57; 47.1%) indicated four categories of information.

Table 6. The Number of Categories of Information Provided in the Reviewed Studies.

Number of Categories Reported	Number of Studies	%
Three	24	19.8%
Four	57	47.1%
Five	30	24.8%
Six	10	8.3%
Total	121	100.0%

We subsequently examined the number of studies that reported full information within each category, the results of which are presented in Figure 7. What stands out in the table is that none of the studies in the sample reported all necessary items. Across the categories, the software used in the L2 eye-tracking studies was specified in 54 studies (44.6%), followed by visual stimuli (*n* = 19; 15.7%). Surprisingly, no study reported all 10 items in the category of "study context and participant demographics", and only one study presented information concerning the five items under "data pre-processing procedures" (see Appendix I for a full presentation of the reporting practices).

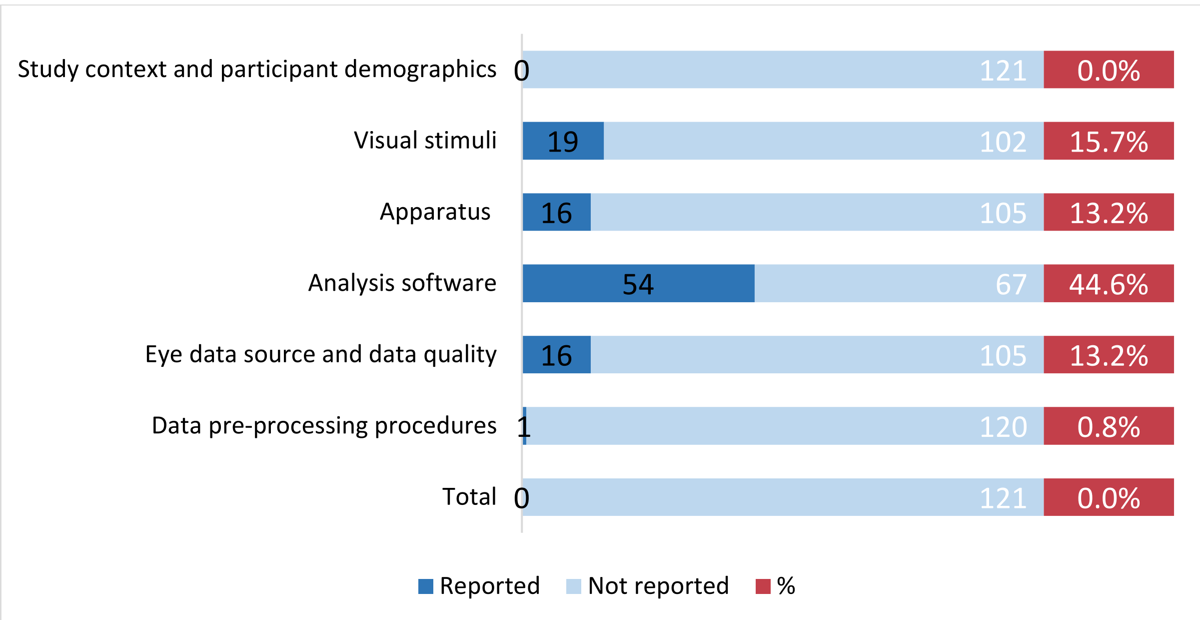


Figure 7. Complete Reporting across Categories.

Finally, Table 7 illustrates the replicability index which is a function of the number of variables reported in each study over the total number of variables necessary for replicability purposes. The number of variables expected to be reported by each study varies slightly based on the modality of the visual stimuli used in the study. Specifically, text-based studies consist of 30 variables and image-based studies include 28 variables, while the number of variables for studies using both text and image stimuli is 31 variables.

Table 7. Replicability Index.

The Percentage of Variables Reported in Each Study	Number of Studies	%
Below 50.0%	43	35.5%
50.0% to 70.0%	73	60.3%
Above 70.0%	5	4.1%

It can be seen that only five studies (4.1%) reported more than 70.0% of the variables, while most studies (*n* = 73; 60.3%) provided between 50.0% and 70.0% of the variables. Notably, 43 studies (35.5%) specified less than 50.0% of the items in their articles, underscoring the prevalent issue of incomplete reporting in previous eye-tracking studies in L2 research. The six groups of variables are unpacked in the following sections.

8. Discussion

The present study synthesized the publications involving eye-tracking in L2 research (*N* = 111). Below, we will discuss our research findings and their implications for L2 eye-tracking research.

8.1. Research Question 1: Areas of Application of Eye-Tracking

The first research question aimed to investigate the types of eye-tracking measures that had to date been used in different research areas. Firstly, eight major research areas emerged together with one mixed area, showing that eye-tracking has been widely adopted as a tool to explore various language skills and linguistic components by L2 researchers. However, the areas appear not to have been investigated with the same depth and thus, certain areas remain under-researched.

Grammar and vocabulary are the most widely investigated areas of research, resonating with the findings of Godfroid (2019). The results indicate that there has been a high level of interest in applying eye-tracking to study L2 processing at the local level (e.g., word level and sentence level). This pattern may be attributed to two main factors. Firstly, experimental designs in word and sentence processing research are vast and varied: there are a myriad of experimental paradigms, input modes and different language learning stages that could be incorporated with eye-tracking to inform lexico-grammatical processing and learning (Godfroid 2019; Keating and Jegerski 2015). In terms of experimental paradigms for examining grammatical processing, there are four types of design: anomaly detection, ambiguity resolution, syntactic dependency formation, and referential processing (Godfroid 2019; Keating and Jegerski 2015), all of which have been adopted across the L2 studies reviewed here. Furthermore, processing and/or learning grammar and vocabulary have been examined through multiple input modes, including reading, listening, reading-while-listening, and viewing, thus allowing for a considerable flexibility in research design. Eye-tracking has also been utilized to investigate both the learning process (e.g., Montero Perez et al. 2015; Winke 2013) and the learning outcome (e.g., Pellicer-Sánchez et al. 2021a; Wong and Ito 2018). This has significantly extended the applicability of eye-tracking across a variety of research designs.

The second reason for the widespread use of eye-tracking in grammar and vocabulary research is the popularity of studies on the effect of attention on language learning and input processing in L2 research (Indrarathne and Kormos 2017; Issa and Morgan-Short 2019). Eye-tracking can capture the attention paid by learners to various linguistic features with an unprecedented level of precision in a natural and non-invasive way. A growing body of eye-tracking research has emerged in the domains of grammar and vocabulary acquisition and instruction, which employs eye-tracking to examine not only the locus but also the amount of attention paid to various target features. This has enabled the eye-tracking method to function as a viable substitute for traditional and less natural methods such as underlining, think-aloud protocols, and mouse-clicking (Duchowski 2017; Godfroid et al. 2013).

It has also been shown that the number of studies on L2 reading and listening has demonstrated a steady increase in recent years. This may be attributed to eye-tracking's potential to shed light on how L2 learners process written and auditory input in real-time within naturalistic settings (Dussias 2010; Roberts and Siyanova-Chanturia 2013). Comparatively, eye-tracking has not achieved widespread adoption in the study of L2 writing. This may be attributed to the ongoing exploration by writing researchers regarding the applicability of eye-tracking in studying the writing process, and only recently have some studies aimed to illustrate the affordance of eye-tracking in this area (e.g., Ranalli et al. 2019). The analysis of participants' eye movements during the composition processes commonly combines a digital screen recording of the process with visualizations of eye movements (Révész et al. 2019). Under these circumstances, the resultant video streams need to be manually annotated by creating moment-by-moment segments. This is a time-consuming procedure, and this methodological complexity can limit the development of L2 eye-tracking writing research. Fixing the screen areas where participants can type in their texts might be offered as a solution, but this can also diminish the authenticity of experiments. It is particularly the case that, in source-based academic writing—wherein authors frequently alternate between texts and/or scroll up and down in texts—an array of cognitive processes such as reading, viewing, skimming, scanning, confusion, confirmation, rebutting and so on are involved which would be extremely difficult, if not impossible, to disentangle in the gaze behavior data. In such cases, writing researchers need to utilize several methods to tap into these processes; for example, Révész et al. (2019) combined keystroke logs, eye-tracking and stimulated recall comment, demonstrating the methodological complexity required in this line of inquiry.

It was also found that the amount of eye-tracking research on validity, speaking and phonology is very limited. In validity research, eye-tracking has emerged as a useful means for validating language tasks or evaluating the construct validity of readability formulas (Nahatame 2021; Révész et al. 2014), illustrating new directions regarding the use of this method. While research on eye-tracking from psychology has provided basic knowledge and assumptions about the application of this method, L2 studies have fundamentally explored phenomena different from other fields, which entails unique experimental designs and features. Thus, it is suggested that more research should be conducted to establish the validity of eye-tracking in L2 research and advance field-specific methodological knowledge. With respect to speaking and phonology, only two studies were identified in each area (speaking, Flecken et al. 2015; Lee and Winke 2018; phonology, Esteve-Gibert and Muñoz 2021; Stone et al. 2018), demonstrating a relative lack of attention to these domains in L2 eye-tracking research and that more scholarly attention should be paid to them. In all, a wealth of research opportunities remains for future studies where the analysis of gaze behaviors can be informative.

8.2. Eye-Tracking Measures Adopted

As discussed earlier, we adapted Lai et al.'s (2013) and Godfroid's (2019) classifications of eye-tracking measures to identify the measures used in our sample. It is evident from the results that there is an uneven distribution of the types of eye-tracking measures used in L2 research. While a substantial portion of studies in the sample adopted fixation measures, many other measure types were used far less frequently, such as regressions, saccades, blinks, and pupil measures.

The wide application of fixations in a range of subareas of L2 studies may be attributed to the fact that they can reveal much about various cognitive operations and online language processing, such as attention (e.g., Godfroid et al. 2013; Issa and Morgan-Short 2019), reading processes (e.g., Traxler et al. 2021), and listening processes (e.g., Kim and Grüter 2021; Mitsugi 2017). Similar to previous studies in general educational research (e.g., Lai et al. 2013), the results of this study demonstrate that fixation temporal measures have been predominantly applied by L2 researchers, followed by fixation count measures. Specifically, the two research designs that have been widely adopted—reading and the visual world paradigm—mainly focused on the analysis of fixation temporal and count measures, respectively (Godfroid 2019; Holmqvist et al. 2011). We found that the distribution of measure types in the field varies based on the topics under investigation. Again, in topics investigated more often through reading (e.g., grammatical processing and reading behaviors), fixation temporal measures are dominant. By contrast, in areas wherein the visual world paradigm is frequently applied (e.g., predictive language processing), fixation count measures such as the proportion of fixations are commonly used. Therefore, the choice of eye-tracking measures should be informed by research questions, research designs, and theories. In other words, there is no one-size-fits-all paradigm in designing eye-tracking research.

A few types of eye-tracking measures were used in a limited number of L2 studies. In particular, saccades, one of the fundamental measures of eye movements, were one of the least frequently used measure indices alongside eye blinks and pupil measures in L2 studies. However, there is one positive trend showing an increase in the application of saccadic eye-movement measures (e.g., Hung et al. 2020; Pellicer-Sánchez et al. 2021b). This is in contrast to the findings of Godfroid (2019) who found no saccadic measures used in her sample of publications (number of papers reviewed = 84). The increasing use of saccades may be attributed to researchers' recognition that saccades can reveal participants' processing of multimodal input (Alemdag and Cagiltay 2018). The results also demonstrate that L2 researchers have expanded the affordances and span of eye-tracking to issues that are commonly addressed using traditional research designs such as interviews and think-aloud protocols.

Combining different gaze behavior measures can better inform different aspects of L2 processing. Nahatame (2021) noted that the information on the global characteristics of gaze patterns during reading is typically indicated by a combination of fixation duration, saccade length, skipping rate and regression rate. However, this pattern is primarily based on findings from L1 reading studies, while L2 research on readers' global reading behavior remains limited. The application of different measure types has the potential to extend the scope of current research beyond its current fronts in investigating L2 processing and comprehension at the local level (e.g., word level and clause level).

So far, in L2 studies, many eye-tracking measures have been underutilized; this is particularly the case for pupil and blink measures which none of the reviewed studies have employed. As shown in the study by Schmidtke (2018), pupil size can be applied in auditory and orthographic language processing and speech production research. Although less discussed in L2 research, blink rate can reveal processes underlying learning and goal-oriented behavior (Eckstein et al. 2017) or the mental workload (Holmqvist et al. 2011; Perkhofer and Lehner 2019). Overall, more research is needed to shine a light on whether and how the underutilized gaze measures can provide L2 researchers with new evidence.

8.3. Research Question Two

We found that L2 researchers investigated three types of cognitive mechanisms through eye-tracking: attention, higher cognitive processes, and cognitive load.

8.3.1. Attention

As the results show, a large number of the L2 studies (n = 51) utilized eye-tracking to gauge participants' attention to various visual stimuli, indicating that attention is a central area of application of eye-tracking in L2 research in terms of cognitive mechanisms. This is attributed to the fact that gaze behavior is a more direct, continuous, and objective measure of overt visual attention compared with other available research techniques (Duchowski 2017; Issa and Morgan-Short 2019). Notably, eye-tracking provides researchers with not only the concurrent distribution or spatial information of attention but also the amount of attention paid or temporal distribution, thus offering a unique quantitative and continuous measure for attention, which cannot be obtained through other research methods such as note-taking or mouse-clicking (Duchowski 2017; Godfroid et al. 2013; Son et al. 2021).

With respect to eye-tracking measures of attention, studies included in this review demonstrated a strong preference for adopting temporal and count measures. It indicates that eye-tracking was used to measure the amount of attention allocated to certain areas (e.g., specific linguistic features in the written input, image and word areas in the multimodal input), rather than the sequence and direction of participants' allocation of visual attention as can be reflected in spatial scale measures. The advantage of measuring attention with fixation and dwell measures using the temporal and count scales is that the interpretations can be straightforward: a higher number of fixations/dwells and longer fixation/dwell durations typically indicate a larger amount of attention (e.g., Alhazmi et al. 2019; Bax 2013; Indrarathne and Kormos 2017). Most studies indeed treat longer fixation time as an indication of higher levels of attention (e.g., Son et al. 2021; Warren et al. 2018). Similarly, higher numbers of fixation counts (e.g., Lee and Jung 2021; Pellicer-Sánchez et al. 2020), longer dwell durations (e.g., Batty 2021; Bax and Chan 2019), and higher numbers of dwell counts (e.g., Bax and Chan 2019; Lee and Révész 2020) are viewed as indicators of higher levels of attention.

There are some other types of gaze behaviors that have been adopted to measure attention, although less frequently, including gaze patterns (e.g., Hung et al. 2020), skipping (e.g., Lee and Révész 2020), regression counts (e.g., Montero Perez 2019), and saccade spatial measures (e.g., Hung et al. 2020). Gaze patterns integrate both spatial and temporal aspects of gaze behaviors, thus providing useful information to draw inferences on participants' patterns of attention (Rahal and Fiedler 2019). For example, Hung et al. (2020) visualized eye movements to represent attention, reporting that fixation-time-based heat maps provided a holistic view regarding L2 readers' visual attention to science text with visuals. Nevertheless, this approach is relatively uncommon, possibly because of the time-consuming nature of manually inspecting the gaze patterns of every participant, likely resulting in bias.

Critically, it is important to realize that eye-tracking is restricted to monitoring foveal vision, while humans can process visual stimuli with their parafoveal vision but cannot be recorded by the eye tracker (Godfroid 2019; Henderson 2003). It is on this basis that words and AOIs that are processed parafoveally may be skipped (Godfroid and Hui 2020). Accordingly, it would be beneficial to measure the duration of fixations on words and AOIs that are fixated by participants, as fixations provide positive evidence of attention (Godfroid and Hui 2020).

In summary, it may be said that there is no fixed measure in eye-tracking to investigate attention across different research designs. The data suggests that researchers can employ various types of eye-tracking measures to gain more insights into attention allocation (see Orquin and Holmqvist 2018, for a recent example). Thus, researchers investigating attention are recommended to consider adopting multiple measures in representing attention, such as fixation temporal and count, dwell temporal and count, and gaze pattern. This will help researchers to better delineate the time course of attention and its oscillation in time and space.

8.3.2. Higher Cognitive Processes

Higher cognitive processes were less frequently examined than attention in L2 eye-tracking studies, likely because it is challenging to properly and confidently link gaze behaviors to hypothesized undergirding cognitive processes (Strohmaier et al. 2020), while there is a more direct and perhaps solid relation between overt visual attention and eye movements (Godfroid 2019; Holmqvist et al. 2011).

A broader range of measure types was applied to explore higher-level cognitive processing. The most frequently used measure type is fixation count followed by fixation temporal measures. This is attributable to the frequent application of the visual world paradigm, in which the proportion of fixation is assessed to infer language comprehension processes (Dussias 2010; Huettig et al. 2011). On the other hand, fixation temporal measures are commonly used in reading-based studies to examine the moment-by-moment comprehension processes underlying reading (e.g., Keating 2009; Traxler et al. 2021). Using fixation temporal measures to gauge cognitive processing in L2 offers several advantages. First, research shows that fixation duration can be reliably used as a proxy for representing the depth of processing (Son et al. 2021). During L2 reading, longer fixations can be indicators of deeper levels of processing (Son et al. 2021), although they could also indicate difficulty in comprehension. Second, eye movements during reading can be carved up into early (e.g., first fixation duration) and late (e.g., total fixation duration) measures. It enables researchers to use a range of early and late measures to uncover the temporal dynamics of processing (Godfroid and Hui 2020).

Regression is also a relatively common measure to infer higher cognitive processes during L2 reading (e.g., Elgort et al. 2018; Fujita and Cunnings 2021; Keating 2009). A notable advantage of using eye-tracking over self-paced reading to examine the reading process is that, reading with eye-tracking allows for capturing and measuring regressions (Godfroid 2019). Regression is considered to represent reanalysis and reflect processing difficulties (Rayner 2009). For example, regression path duration is often interpreted as the time required to resolve a processing challenge (Godfroid 2019). In terms of the eye-tracking results, higher regression rates or longer regressions can be representative of difficulties in lexical access or text comprehension (Elgort et al. 2018; Keating 2009). Based on such understanding, measures of regressive eye movements have been utilized to investigate the process of, for example, anomaly detection during online sentence comprehension (Keating 2009) and semantic integration in contextual word learning from reading (Elgort et al. 2018). However, it seems that our knowledge of regressions in L2 processing is insufficient, which may impede the operationalization and interpretation of this measure in L2 research. For instance, it is possible that people regress to and refixate on a previous part because that word or AOI is of high relevance to the task (Orquin and Holmqvist 2018). Nonetheless, this interpretation is rarely mentioned in the L2 literature compared to comprehension difficulties or semantic integration as possible causes of regressions (e.g., Roberts and Siyanova-Chanturia 2013). A possible reason for the limited understanding of regressions in L2 processing is that "regressions are not particularly well understood because it is difficult to control them experimentally" (Rayner 2009, p. 1460). Overall, the absence of a reliable understanding of regressions may jeopardize the validity of inferences drawn from regressive eye movements.

In sum, L2 researchers have leveraged numerous eye-tracking measures to tap different cognitive processes. Nevertheless, using eye-tracking to draw inferences concerning higher cognitive processes can be challenging because of the simultaneous effects of multiple unobservable cognitive processes on eye movements. The same measure has been found to be interpreted in different ways (see, e.g., Bax 2015; Meghanathan et al. 2015). Moreover, two people could fixate on an AOI for the same amount of time but for different reasons. Take the fixation duration as an example; this metric can be associated with both the number of distractors that enter the mind and the amount of attention deployed to process the target stimuli (Meghanathan et al. 2015), which can pose a problem to the interpretation of fixation in multimodal reading if fixation duration is taken as a proxy for integration and comprehension. Even in unimodal environments, the interpretation of fixation duration is not straightforward and tightly controlled experimental designs are needed to draw accurate inferences from fixation duration and other metrics. As L2 researchers are exploiting the new possibility of eye-tracking in the studies of L2 processing, greater efforts need to be devoted to disentangling different processes underlying eye movements.

8.3.3. Cognitive Load

Cognitive load refers to "the load that performing a particular task imposed on the participant's cognitive system" (Paas et al. 2003, p. 64). A smaller number of L2 studies (n = 3) have probed cognitive load using eye-tracking, even though cognitive load can be more easily derived from gaze behaviors (Meghanathan et al. 2015). This construct is considerably under-investigated in the L2 literature, too, and there is much room for development and innovation in this line of research. Although only three studies examining cognitive load have been found, a variety of measure types have been used. Fixation temporal and count measures have been utilized more frequently, and a higher fixation rate or a longer fixation duration indicates a greater cognitive load (Aryadoust et al. 2022; Révész et al. 2014). The other measure types have also proven to be useful in measuring cognitive load, although their application has been far and few between, such as dwell temporal, dwell count, skip, regression count and saccade spatial measures. The choice of eye-tracking measure and the interpretation of the collected metrics will depend on the type of the task used. In a study about measuring cognitive load imposed by test methods, Aryadoust et al. (2022) used dwell temporal and dwell count measures, wherein lower visit rates and higher normalized visit duration were treated as indications of lower cognitive load in the while-listening-performance tests.

Another useful measure of cognitive load is pupil diameter, whose relationship with cognitive load was discovered several decades ago by Hess and Polt (1964). Following this discovery, pupil size has been widely used in a variety of research designs (see van der Wel and van Steenbergen 2018), although being underappreciated in L2 research. Task-evoked pupil dilation has been found to be sensitive to cognitive load (Beatty and Lucero-Wagoner 2000), such that the size of the pupil enlarges proportionally in states of high mental load (Perkhofer and Lehner 2019).

Nevertheless, we note that using pupil diameter as a measure of cognitive load is not without limitations. The cognitive effects on pupil diameters are small but changes in pupil dilation can be triggered by a variety of extraneous factors, such as luminance and off axis-distortion (Holmqvist et al. 2011; Krejtz et al. 2018). Small changes in pupil diameter can easily be drowned in the large changes due to variations in light intensity, and thus the use of this measure will require the implementation of a tight experimental setup and design (Holmqvist et al. 2011).

8.4. Research Question 3: Replicability of L2 Eye-Tracking Studies

The third research question sought to provide a survey of the transparent and replicable reporting practices of L2 eye-tracking studies. We developed a list of 33 items critical to replicating an L2 eye-tracking study, which was informed by previous eye-tracking reporting guidelines (Carter and Luke 2020; Fiedler et al. 2020; King et al. 2019), methodological handbooks (Godfroid 2019; Holmqvist et al. 2011) and several empirical L2 eye-tracking studies (e.g., Plonsky and Gonulal 2015; Traxler et al. 2021).

The results revealed that the majority of the reviewed studies did typically not specify a number of categories and features of their studies. Taken together, the results from different analyses largely converged on the fact that L2 eye-tracking studies are not sufficiently transparent and complete in their reporting practices. Similar findings have been documented in previous reviews evaluating the reporting practices of eye-tracking in other fields, such as mathematics research (Strohmaier et al. 2020), communication science (King et al. 2019), and behavioral decision-making (Fiedler et al. 2020). Relatedly, the results also mirror those of the previous reviews examining the reporting of statistical methods or research tools in L2 research (Crowther et al. 2021; Plonsky and Gonulal 2015), where much of the critical information of each method was overlooked in the articles.

This worrying finding may be attributed to two main reasons. First, L2 eye-tracking studies have not established an agreed-upon reporting guidelines for L2 researchers to follow in the past years. Due to different considerations with regard to the nature of the respective research fields, reporting guidelines are also often biased towards specific research fields, and most of the prior reporting guidelines of eye-tracking are incomplete and inconsistent regarding what information to report (Holmqvist et al. 2023). Second, journal articles are often limited in length, and thus some of the information cannot be sufficiently specified in publications.

Eye-tracking studies in this review have shown a high level of variability and flexibility in methodological aspects regarding the apparatus, the analysis software, procedures and the parameters used. With such high degrees of freedom, eye-tracking researchers should make an effort to ensure the replicability of their studies by transparently reporting their research. This effort would contribute to the reliability and generalizability of the results obtained from the eye-tracking method and the scientific progress of the field. It should be noted that this study has explored replicability from the perspective of the extent to which L2 eye-tracking studies could be performed again by other researchers following the information and procedures documented in the original publications. The reasons for why a study cannot be replicated are complex and many, but transparent reporting with a high level of details from study design, sampling, choices of hardware and software, techniques and parameters used to process data will no doubt enhance replicability.

9. Conclusions

This systematic review sought to scope the application of eye-tracking in L2 research and examine the extent to which L2 eye-tracking studies are replicable from the perspective of reporting practices. The findings indicate a growing adoption of eye-tracking in L2 research. Eye-tracking was frequently used in grammar and vocabulary research; fixation temporal and count measures were the most frequently used measures; and three cognitive mechanisms were investigated, as previously discussed. However, our review also highlights that eye-tracking is just beginning to emerge in many areas, such as L2 writing and speaking. Furthermore, our classification of eye-tracking measures enabled us to demonstrate the limited utilization of available eye-tracking measures across various domains of L2 research. We suggest that future researchers explore these under-researched areas as well as cognitive mechanisms to attain a comprehensive understanding of L2 learners' gaze behaviors during language processing and learning. Finally, the evidence of insufficient transparency emphasizes the need for more detailed reporting practices in future eye-tracking studies.

This review is subject to several limitations. One issue with the current review is that the studies reviewed here may not represent the entire L2 eye-tracking literature, since the data were chosen from tier-1 (quartile-1) English journals in the field. We suggest that future researchers should investigate articles published in other journals or sources (e.g., tier-2, tier-3, and even tier-4 journals), so that a more comprehensive review of the application and uses of eye-tracking in the field will be generated. In addition, future researchers should consider reviewing the articles that are published in bilingualism and multilingualism journals. It should also be noted that this study did not extract information from every possible source relevant to each reviewed study, which may impact the replicability index of L2 eye-tracking studies. Due to word limit requirements imposed by journals, researchers may need to selectively report information while potentially providing supplementary details through other channels. For instance, with the promotion of open research practices, researchers are making their materials and data accessible through repositories like IRIS (Instruments and Data for Research in Language Studies) (Marsden and Morgan-Short 2023). Nevertheless, our list of variables for replicability purposes remains critical, and the absence of such information in the paper could signify reduced replicability (Derrick 2016). We hope that the results of this study will encourage better design and reporting practices in future eye-tracking research.

Author Contributions: Conceptualization, X.H. and V.A.; methodology, X.H. and V.A.; software, X.H.; validation, X.H. and V.A.; formal analysis, X.H. and V.A.; investigation, X.H. and V.A.; resources, X.H. and V.A.; data curation, X.H.; writing—original draft preparation, X.H.; writing—review and editing, X.H. and V.A.; visualization, X.H.; supervision, V.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Examples and Definitions of Commonly Used Eye-Tracking Measures (Informed by Godfroid 2019; Holmqvist et al. 2011; Lai et al. 2013)

Measure	Scale	Definition or Calculation
Fixation		
Time to first fixation	Temporal	The time period from entering the AOI until the first fixation is made.
First fixation duration	Temporal	The duration of the initial fixation on an AOI.
Gaze duration	Temporal	The sum of all fixations recorded for a single-word interest area until the eyes move away from the area.
First pass reading time	Temporal	The sum of all fixations recorded for a multi-word interest area until the eyes move away from the area.
Second pass time	Temporal	The aggregated duration of all fixations made within an interest area during the second visit to the area, including instances where the AOI was initially skipped.
Rereading time	Temporal	The aggregated duration of all fixations in an AOI, excluding those made during the first pass.
Average fixation duration	Temporal	The mean of fixation durations on each AOI.
Total fixation duration	Temporal	The aggregated duration of all fixations in an AOI.
Number of fixations	Count	The number of fixations made in an AOI.
Proportion of fixations	Count	The proportion of total fixations that are directed to an AOI, or the number of fixations between AOIs and between experimental groups.
Fixation position	Spatial	The location of a fixation.
Saccade		
Saccade duration	Temporal	The amount of time that the eyes take to move between two fixations.
Saccade count	Count	The number of saccades made in an AOI or in a trial.
Saccade length/amplitude	Spatial	The distance between two consecutive fixations.
Dwell		
Dwell time	Temporal	The amount of time that the eyes spend in an AOI during a dwell, including the durations of fixations and non-fixations.
Total reading time	Temporal	Total time spent within an AOI or spent on a reading task.
Total visit duration	Temporal	The aggregated duration of all visits to a specific AOI.
Total number of visits	Count	The total count of visits to a specified AOI.
Dwell rate	Count	The number of entries into a specific area of interest per minute.
Regression		
Regression path duration/go-past time	Temporal	The duration from the first entry in an AOI until exiting that AOI in the reading direction.
Regression rate	Count	The number of regressions per unit (e.g., second, line, paragraph).
Regression in	Count	A backward eye movement that falls on a selected AOI.
Regression out	Count	A backward eye movement that originates from a selected AOI.

Measure	Scale	Definition or Calculation
Skip		
Skipping proportion/rate	Count	The proportion of participants who never fixate on a selected AOI.
Skip count	Count	The total count of instances where AOI is passed over.
Pupil		
Pupil diameter	Spatial	The pupil size for the current position of the eye.
Pupil dilation latency	Temporal	The time period from the onset of a stimulus until the beginning of pupil dilation.
Blink		
Blink rate	Count	The number of blinks per unit of time.
Blink duration	Temporal	The time period from the moving down of the eyelid until it opens up completely.
Gaze pattern		
Heatmap	NA	A visual representation of the distribution of participants' eye movements across a screen, using a range of warm and cold colors.
Scanpath	Spatial	A visual or numerical representation of the trace of fixations and saccades.

Appendix B. List of Journals

Applied Linguistics
Applied Psycholinguistics
Assessing Writing
Computer Assisted Language Learning
Journal of Second Language Writing
Language Learning
Language Learning and Development
Language Learning and Technology
Language Learning Journal
Language Teaching
Language Teaching Research
Language Testing
Modern Language Journal
RELC Journal
Studies in Second Language Acquisition
Studies in Second Language Learning and Teaching
System
TESOL Quarterly

Appendix C. Scopus Search Code

(SRCTITLE ("Applied Linguistics") OR SRCTITLE ("Modern Language Journal") OR SRCTITLE ("Language Learning") OR SRCTITLE ("Language Testing") OR SRCTITLE ("Studies in Second Language Acquisition") OR SRCTITLE ("Journal of Second Language Writing") OR SRCTITLE ("Language Teaching") OR SRCTITLE ("TESOL Quarterly") OR SRCTITLE ("Language Teaching Research") OR SRCTITLE ("Computer Assisted Language Learning") OR SRCTITLE ("Language Learning and Technology") OR SRCTITLE ("Annual Review of Linguistics") OR SRCTITLE ("System") OR SRCTITLE ("Assessing Writing") OR SRCTITLE ("Studies in Second Language Learning and Teaching") OR SRCTITLE ("English for Specific Purposes") OR SRCTITLE ("Language Awareness") OR SRCTITLE ("ReCALL") OR SRCTITLE ("Applied Psycholinguistics") OR SRCTITLE ("ELT Journal") OR SRCTITLE ("RELC Journal")) AND (TITLE-ABS-KEY ("eye tracking") OR TITLE-ABS-KEY ("eye movement")) AND (LIMIT-TO (SUBJAREA, "SOCI") OR LIMIT-TO (SUBJAREA, "ARTS") OR LIMIT-TO (SUBJAREA, "PSYC")) AND (LIMIT-TO (EXACT-SRCTITLE, "Studies In Second Language Acquisition") OR LIMIT-TO (EXACTSRCTITLE, "Applied Psycholinguistics") OR LIMIT-TO (EXACTSRCTITLE, "Language Learning") OR LIMIT-TO (EXACTSRCTITLE, "System") OR LIMIT-TO (EXACTSRCTITLE, "Computer Assisted Language Learning") OR LIMIT-TO (EXACTSRCTITLE, "Language Testing") OR LIMIT-TO (EXACTSRCTITLE, "Modern Language Journal") OR LIMIT-TO (EXACTSRCTITLE, "Language Teaching Research") OR LIMIT-TO (EXACTSRCTITLE, "Language Learning And Development") OR LIMIT-TO (EXACTSRCTITLE, "Applied Linguistics") OR LIMIT-TO (EXACTSRCTITLE, "Language Learning And Technology") OR LIMIT-TO (EXACTSRCTITLE, "Assessing Writing") OR LIMIT-TO (EXACTSRCTITLE, "Relc Journal") OR LIMIT-TO (EXACTSRCTITLE, "TESOL Quarterly") OR LIMIT-TO (EXACTSRCTITLE, "Journal Of Second Language Writing") OR LIMIT-TO (EXACTSRCTITLE, "Language Learning Journal") OR LIMIT-TO (EXACTSRCTITLE, "Language Teaching") OR LIMIT-TO (EXACTSRCTITLE, "Studies In Second Language Learning And Teaching")) AND (LIMIT-TO (LANGUAGE, "English"))

Appendix D. List of Publications Included in the Systematic Review

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
1	Nahatame S.	Text readability and processing effort in second language reading: a computational and eye-tracking investigation	2021	Language Learning	71	4	1004–1043	10.1111/lang.12455
2	Batty A.O.	An eye-tracking study of attention to visual cues in L2 listening tests	2021	Language Testing	38	4	511–535	10.1177/0265532220951504
3	Lozano-Argüelles C., Sagarra N.	Interpreting experience enhances the use of lexical stress and syllabic structure to predict L2 word endings	2021	Applied Psycholinguistics	42	5	1135–1157	10.1017/S0142716421000217
4	Ge H., Mulders I., Kang X., Chen A., Yip V.	Processing focus in native and non-native speakers of English: an eye-tracking study in the visual world paradigm	2021	Applied Psycholinguistics	42	4	1057–1088	10.1017/S0142716421000230
5	Prichard C., Atkins A.	Evaluating the vocabulary coping strategies of L2 readers: an eye tracking study	2021	TESOL Quarterly	55	2	593–620	10.1002/tesq.3005
6	Traxler M.J., Banh T., Craft M.M., Winsler K., Brothers T.A., Hoversten L.J., Piñar P., Corina D.P.	Word skipping in deaf and hearing bilinguals: Cognitive control over eye movements remains with an increased perceptual span	2021	Applied Psycholinguistics	42	3	601–630	10.1017/S0142716420000740

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
7	Wiener S., Ito K., Speer S.R.	Effects of multitalker input and instructional methods on the dimension-based statistical learning of syllable-tone combinations	2021	Studies in Second Language Acquisition	43	1	155–180	10.1017/S0272263120000418
8	Kim H., Grüter T.	Predictive processing of implicit causality in a second language	2021	Studies in Second Language Acquisition	43	1	133–154	10.1017/S0272263120000443
9	Fujita H., Cunnings I.	Lingering misinterpretation in native and nonnative sentence processing: evidence from structural priming	2021	Applied Psycholinguistics	42	2	475–504	10.1017/S0142716420000351
10	Pellicer-Sánchez A., Conklin K., Vilkaitė-Lozdienė L.	The effect of pre-reading instruction on vocabulary learning: an investigation of L1 and L2 readers' eye movements	2021a	Language Learning	71	1	162–203	10.1111/lang.12430
11	Pellicer-Sánchez A., Conklin K., Rodgers M.P., Parente F.	The effect of auditory input on multimodal reading comprehension: an examination of adult readers' eye movements	2021b	Modern Language Journal	105	4	936–956	10.1111/modl.12743
12	Maie R., Godfroid A.	Controlled and automatic processing in the acceptability judgment task: an eye-tracking study	2021	Language Learning	72	1	158–197	10.1111/lang.12474
13	Gánem-Gutiérrez G.A., Gilmore A.	A mixed methods case study on the use and impact of web-based lexicographic tools on L2 writing	2021	Computer Assisted Language Learning			1–24	10.1080/09588221.2021.1987273
14	Son M., Lee J., Godfroid A.	Attention to form and meaning revisited	2021	Studies in Second Language Acquisition	44	3	788–817	10.1017/S0272263121000565
15	Freeman M.R., Marian V.	Visual word recognition in bilinguals	2021	Studies in Second Language Acquisition	44	3	759–787	10.1017/S027226312100053X
16	Lipski J.M.	Language revitalization as L2 shadow boxing	2021	Studies in Second Language Acquisition	43	1	220–235	10.1017/S0272263120000339

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
17	Spit S., Andringa S., Rispens J., Aboh E.O.	The effect of explicit instruction on implicit and explicit linguistic knowledge in kindergartners	2021	Language Learning and Development	18	2	201–228	10.1080/15475441.2021.1941968
18	Lee M., Jung J.	Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: a mixed methods study	2021	Language Teaching Research				10.1177/13621688211034640
19	Nisbet K., Bertram R., Erlinghagen C., Pieczykolan A., Kuperman V.	Quantifying the difference in reading fluency between L1 and L2 readers of English	2021	Studies in Second Language Acquisition	44	2	407–434	10.1017/S0272263121000279
20	Aryadoust V., Foo S., Ng L.Y.	What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments?	2021	Language Testing	39	1	56–89	10.1177/02655322211026876
21	Cheng Y., Rothman J., Cunnings I.	Parsing preferences and individual differences in nonnative sentence processing: evidence from eye movements	2021	Applied Psycholinguistics	42	1	129–151	10.1017/S014271642000065X
22	Esteve-Gibert N., Muñoz C.	Preschoolers benefit from a clear sound-referent mapping to acquire nonnative phonology	2021	Applied Psycholinguistics	42	1	77–100	10.1017/S0142716420000600
23	Grüter T., Rohde H.	Limits on expectation-based processing: use of grammatical aspect for co-reference in L2	2021	Applied Psycholinguistics	42	1	51–75	10.1017/S0142716420000582
24	Holzknecht F., McCray G., Eberharter K., Kremmel B., Zehentner M., Spiby R., Dunlea J.	The effect of response order on candidate viewing behaviour and item difficulty in a multiple-choice listening test	2021	Language Testing	38	1	41–61	10.1177/0265532220917316
25	Hung Y.-N., Kuo H.-Y., Liao S.-C.	Seeing what they see: elementary EFL students reading science texts	2020	RELC Journal	51	3	397–411	10.1177/0033688219854475

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
26	THAM I., CHAU M.H., THANG S.M.	Bilinguals' processing of lexical cues in L1 and L2: an eye-tracking study	2020	Computer Assisted Language Learning	33	7	665–687	10.1080/09588221.2019.1588329
27	Wolter B., Yamashita J., Leung C.Y.	Conceptual transfer and lexical development in adjectives of space: evidence from judgments, reaction times, and eye tracking	2020	Applied Psycholinguistics	41	3	595–625	10.1017/S0142716420000107
28	Zhou W., Ye W., Yan M.	Alternating-color words facilitate reading and eye movements among second-language learners of Chinese	2020	Applied Psycholinguistics	41	3	685–699	10.1017/S0142716420000211
29	Benati A.	The effects of structured input and traditional instruction on the acquisition of the English causative passive forms: an eye-tracking study measuring accuracy in responses and processing patterns	2020	Language Teaching Research				10.1177/1362168820928577
30	Kang H., Kweon S.-O., Choi S.	Using eye-tracking to examine the role of first and second language glosses	2020	Language Teaching Research				10.1177/1362168820928567
31	Rusk B.V., Paradis J., Järvikivi J.	Comprehension of English plural-singular marking by Mandarin-L1 and early L2-immersion learners	2020	Applied Psycholinguistics	41	3	579–593	10.1017/S0142716420000089
32	Lee M., Révész A.	Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks	2020	Studies in Second Language Acquisition	42	3	625–651	10.1017/S0272263120000108
33	Pellicer-Sánchez A., Tragant E., Conklin K., Rodgers M., Serrano R., Llanes Á.	Young learners' processing of multimodal input and its impact on reading comprehension	2020	Studies in Second Language Acquisition	42	3	577–598	10.1017/S0272263120000091

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
34	Tomasuolo E., Roccaforte M., Di Fabio A.	Reading and deafness: eye tracking in deaf readers with different linguistic background	2019	Applied Linguistics	40	6	992–1008	10.1093/applin/amy049
35	Koval N.G.	Testing the deficient processing account of the spacing effect in second language vocabulary learning: evidence from eye tracking	2019	Applied Psycholinguistics	40	5	1103–1139	10.1017/S0142716419000158
36	Lee J.F., Doherty S.	Native and nonnative processing of active and passive sentences	2019	Studies in Second Language Acquisition	41	4	853–879	10.1017/S027226311800027X
37	Montero Perez M.	Pre-learning vocabulary before viewing captioned video: an eye-tracking study	2019	Language Learning Journal	47	4	460–478	10.1080/09571736.2019.1638623
38	Chukharev-Hudilainen E., Saricaoglu A., Torrance M., Feng H.-H.	Combined deployable keystroke logging and eyetracking for investigating L2 writing fluency	2019	Studies in Second Language Acquisition	41	3	583–604	10.1017/S027226311900007X
39	Michel M., O'Rourke B.	What drives alignment during text chat with a peer vs. a tutor? Insights from cued interviews and eye-tracking	2019	System	83		50–63	10.1016/j.system.2019.02.009
40	El Ebyary K., Windeatt S.	Eye tracking analysis of EAP Students' regions of interest in computer-based feedback on grammar, usage, mechanics, style and organization and development	2019	System	83		36–49	10.1016/j.system.2019.03.007
41	Bax S., Chan S.	Using eye-tracking research to investigate language test validity and design	2019	System	83		64–78	10.1016/j.system.2019.01.007
42	Révész A., Michel M., Lee M.	Exploring second language writers' pausing and revision behaviors	2019	Studies in Second Language Acquisition	41	3	605–631	10.1017/S027226311900024X
43	Ranalli J., Feng H.-H., Chukharev-Hudilainen E.	The affordances of process-tracing technologies for supporting L2 writing instruction	2019	Language Learning and Technology	23	2	1–11	

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
44	Gass S., Winke P., Isbell D.R., Ahn J.	How captions help people learn languages: a working-memory, eye-tracking study	2019	Language Learning and Technology	23	2	84–104	https://doi.org/10125/44684
45	Ito K., Wong W.	Processing instruction and the effects of input modality and voice familiarity on the acquisition of the French causative construction	2019	Studies in Second Language Acquisition	41	2	443–468	10.1017/S0272263118000281
46	Issa B.I., Morgan-Short K.	Effects of external and internal attentional manipulations on second language grammar development	2019	Studies in Second Language Acquisition	41	2	389–417	10.1017/S027226311800013X
47	Vilkaite L., Schmitt N.	Reading collocations in an L2:Do collocation processing benefits extend to non-adjacent collocations?	2019	Applied Linguistics	40	2	329–354	10.1093/applin/amx030
48	Curcic M., Andringa S., Kuiken F.	The role of awareness and cognitive aptitudes in L2 predictive language processing	2019	Language Learning	69		42–71	10.1111/lang.12321
49	Alhazmi K., Milton J., Johnston S.	Examining 'vowel blindness' among native Arabic speakers reading English words from the perspective of eye-tracking	2019	System	80		235–245	10.1016/j.system.2018.12.005
50	Tragant Mestres E., Pellicer-Sánchez A.	Young EFL learners' processing of multimodal input: examining learners' eye movements	2019	System	80		212–223	10.1016/j.system.2018.12.002
51	Wong K.M., Samudra P.G.	L2 vocabulary learning from educational media: extending dual-coding theory to dual-language learners	2019	Computer Assisted Language Learning	34	8	1182–1204	10.1080/09588221.2019.1666150
52	Warren P., Boers F., Grimshaw G., Siyanova-Chanturia A.	The effect of gloss type on learners' intake of new words during reading: evidence from eye-tracking	2018	Studies in Second Language Acquisition	40	4	883–906	10.1017/S0272263118000177

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
53	Jung J., Révész A.	The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions	2018	Studies in Second Language Acquisition	40	4	755–780	10.1017/S0272263118000165
54	Peters R.E., Grüter T., Borovsky A.	Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals	2018	Applied Psycholinguistics	39	6	1439–1469	10.1017/S0142716418000383
55	Connell K., Hüls S., Martínez-García M.T., Qin Z., Shin S., Yan H., Tremblay A.	English learners' use of segmental and suprasegmental cues to stress in lexical access: an eye-tracking study	2018	Language Learning	68	3	635–668	10.1111/lang.12288
56	Lee M., Révész A.	Promoting grammatical development through textually enhanced captions: An eye-tracking study	2018	Modern Language Journal	102	3	557–577	10.1111/modl.12503
57	Veivo O., Porretta V., Hyönä J., Järvikivi J.	Spoken second language words activate native language orthographic information in late second language learners	2018	Applied Psycholinguistics	39	5	1011–1032	10.1017/S0142716418000103
58	Indrarathne B., Ratajczak M., Kormos J.	Modelling changes in the cognitive processing of grammar in implicit and explicit learning conditions: insights from an eye-tracking study	2018	Language Learning	68	3	669–708	10.1111/lang.12290
59	Prichard C., Atkins A.	L2 readers' global processing and selective attention: an eye tracking study	2018	TESOL Quarterly	52	2	445–456	10.1002/tesq.423
60	Gánem-Gutiérrez G.A., Gilmore A.	Tracking the real-time evolution of a writing event: second language writers at different proficiency levels	2018	Language Learning	68	2	469–506	10.1111/lang.12280

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
61	Wong W., Ito K.	The effects of processing instruction and traditional instruction on L2 online processing of the causative construction in French: an eye-tracking study	2018	Studies in Second Language Acquisition	40	2	241–268	10.1017/S0272263117000274
62	Elgort I., Brysbaert M., Stevens M., Van Assche E.	Contextual word learning during reading in a second language: an eye-movement study	2018	Studies in Second Language Acquisition	40	2	341–366	10.1017/S0272263117000109
63	Mohamed A.A.	Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning	2018	Studies in Second Language Acquisition	40	2	269–293	10.1017/S0272263117000092
64	Stone A., Petitto L.-A., Bosworth R.	Visual sonority modulates infants' attraction to sign language	2018	Language Learning and Development	14	2	130–148	10.1080/15475441.2017.1404468
65	Ranalli J., Feng H.-H., Chukharev-Hudilainen E.	Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing	2018	Assessing Writing	36		77–89	10.1016/j.asw.2018.03.007
66	Lee S., Winke P.	Young learners' response processes when taking computerized tasks for speaking assessment	2018	Language Testing	35	2	239–269	10.1177/0265532217704009
67	Hopp H., Lemmerth N.	Lexical and syntactic congruency in L2 predictive gender processing	2018	Studies in Second Language Acquisition	40	1	171–199	10.1017/S0272263116000437
68	McCray G., Brunfaut T.	Investigating the construct measured by banked gap-fill items: evidence from eye-tracking	2018	Language Testing	35	1	51–73	10.1177/0265532216677105
69	Cunnings I., Fotiadou G., Tsimpli I.	Anaphora resolution and reanalysis during L2 sentence processing	2017	Studies in Second Language Acquisition	39	4	621–652	10.1017/S0272263116000292
70	Boers F., Warren P., Grimshaw G., Siyanova-Chanturia A.	On the benefits of multimodal annotations for vocabulary uptake from reading	2017	Computer Assisted Language Learning	30	7	709–725	10.1080/09588221.2017.1356335

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
71	Indrarathne B., Kormos J.	Attentional processing of input in explicit and implicit conditions	2017	Studies in Second Language Acquisition	39	3	401–430	10.1017/S027226311600019X
72	Muñoz C.	The role of age and proficiency in subtitle reading. an eye-tracking study	2017	System	67		77–86	10.1016/j.system.2017.04.015
73	Mitsugi S.	Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm	2017	Applied Psycholinguistics	38	4	953–983	10.1017/S0142716416000515
74	Choi S.	Processing and learning of enhanced English collocations: An eye movement study	2017	Language Teaching Research	21	3	403–426	10.1177/1362168816653271
75	Carrol G., Conklin K., Gyllstad H.	Found in translation: the influence of the L1 on the reading of idioms in a L2	2016	Studies in Second Language Acquisition	38	3	403–443	10.1017/S0272263115000492
76	Loewen S., Inceoglu S.	The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense	2016	Studies in Second Language Learning and Teaching	6	1	89–110	10.14746/ssllt.2016.6.1.5
77	Godfroid A., Spino L.A.	Reconceptualizing reactivity of thinking alouds and eye tracking: Absence of evidence is not evidence of absence	2015	Language Learning	65	4	896–928	10.1111/lang.12136
78	Suvorov R.	The use of eye tracking in research on video-based second language (L2) listening assessment: a comparison of context videos and content videos	2015	Language Testing	32	4	463–483	10.1177/0265532214562099
79	Winke P., Lim H.	ESL essay raters' cognitive processes in applying the Jacobs et al. rubric: an eye-movement study	2015	Assessing Writing	25		37–53	10.1016/j.asw.2015.05.002
80	Montero Perez M., Peters E., Desmet P.	Enhancing vocabulary learning through captioned video: an eye-tracking study	2015	Modern Language Journal	99	2	308–328	10.1111/modl.12215

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
81	Andringa S., Curcic M.	How explicit knowledge affects online L2 processing: evidence from differential object marking acquisition	2015	Studies in Second Language Acquisition	37	2	237–268	10.1017/S0272263115000017
82	Godfroid A., Loewen S., Jung S., Park J.-H., Gass S., Ellis R.	Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns	2015	Studies in Second Language Acquisition	37	2	269–297	10.1017/S0272263114000850
83	Cintrón-Valentín M., Ellis N.C.	Exploring the interface: explicit focus-on-form instruction and learned attentional biases in L2 Latin	2015	Studies in Second Language Acquisition	37	2	197–235	10.1017/S0272263115000029
84	Alsadoon R., Heift T.	Textual input enhancement for vowel blindness: a study with Arabic ESL learners	2015	Modern Language Journal	99	1	57–79	10.1111/modl.12188
85	Flecken M., Carroll M., Weimar K., Von Stutterheim C.	Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French-German L2 users	2015	Modern Language Journal	99	S1	100–122	10.1111/j.1540-4781.2015.12181.x
86	Pellicer-Sánchez A.	Incidental L2 vocabulary acquisition from and while reading: an eye-tracking study	2014	Studies in Second Language Acquisition	38	1	97–130	10.1017/S0272263115000224
87	Lim J.H., Christianson K.	Second language sensitivity to agreement errors: evidence from eye movements during comprehension and translation	2014	Applied Psycholinguistics	36	6	1283–1315	10.1017/S0142716414000290
88	Kim E., Montrul S., Yoon J.	The on-line processing of binding principles in second language acquisition: evidence from eye- tracking	2014	Applied Psycholinguistics	89	2	1317–1374	10.1017/S0142716414000307

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
89	Ellis N.C., Hafeez K., Martin K.I., Chen L., Boland J., Sagarra N.	An eye-tracking study of learned attention in second language acquisition	2014	Applied Psycholinguistics	35	3	547–579	10.1017/S0142716412000501
90	Bisson M.-J., Van Heuven W.J.B., Conklin K., Tunney R.J.	Processing of native and foreign language subtitles in films: an eye tracking study	2014	Applied Psycholinguistics	35	2	399–418	10.1017/S0142716412000434
91	Révész A., Sachs R., Hama M.	The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts	2014	Language Learning	64	3	615–650	10.1111/lang.12061
92	Liu P.-L.	Using eye tracking to understand the responses of learners to vocabulary learning strategy instruction and use	2014	Computer Assisted Language Learning	27	4	330–343	10.1080/09588221.2014.881383
93	Bax S.	The cognitive processing of candidates during reading tests: evidence from eye-tracking	2013	Language Testing	30	4	441–465	10.1177/0265532212473244
94	Shintani N., Ellis R.	The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article	2013	Journal of Second Language Writing	22	3	286–306	10.1016/j.jslw.2013.03.011
95	Van Assche E., Duyck W., Brysbaert M.	Verb processing by bilinguals in sentence contexts	2013	Studies in Second Language Acquisition	35	2	237–259	10.1017/S0272263112000873
96	Dussias P.E., Valdés Kroff J.R., Guzzardo Tamargo R.E., Gerfen C.	When gender and looking go hand in hand	2013	Studies in Second Language Acquisition	35	2	353–387	10.1017/S0272263112000915
97	Spinner P., Gass S.M., Behney J.	Ecological validity in eye-tracking	2013	Studies in Second Language Acquisition	35	2	389–415	10.1017/S0272263112000927
98	Godfroid A., Uggen M.S.	Attention to irregular verbs by beginning learners of German	2013	Studies in Second Language Acquisition	35	2	291–322	10.1017/S0272263112000897

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
99	Sagarra N., Ellis N.C.	From seeing adverbs to seeing verbal morphology	2013	Studies in Second Language Acquisition	35	2	261–290	10.1017/S0272263112000885
100	Winke P.M.	The effects of input enhancement on grammar learning and comprehension	2013	Studies in Second Language Acquisition	35	2	323–352	10.1017/S0272263112000903
101	Winke P., Gass S., Sydorenko T.	Factors influencing the use of captions by foreign language learners: an eye-tracking study	2013	Modern Language Journal	97	1	254–275	10.1111/j.1540-4781.2013.01432.x
102	Godfroid A., Boers F., Housen A.	An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking	2013	Studies in Second Language Acquisition	35	3	483–517	10.1017/S0272263113000119
103	Smith B.	Eye tracking as a measure of noticing: a study of explicit recasts in SCMC	2012	Language Learning and Technology	16	3	53–81	http://dx.doi.org/10125/44300
104	Felser C., Cunnings I.	Processing reflexives in a second language: the timing of structural and discourse-level constraints	2012	Applied Psycholinguistics	33	3	571–603	10.1017/S0142716411000488
105	Felser C., Cunnings I., Batterham C., Clahsen H.	The timing of island effects in nonnative sentence processing	2012	Studies in Second Language Acquisition	34	1	67–98	10.1017/S0272263111000507
106	Bolger P.A., Zapata G.	Semantic categories and context in L2 vocabulary learning	2011	Language Learning	61	2	614–646	10.1111/j.1467-9922.2010.00624.x
107	Keating G.D.	Sensitivity to violations of gender agreement in native and nonnative Spanish: an eye-movement investigation	2009	Language Learning	59	3	503–535	10.1111/j.1467-9922.2009.00516.x
108	Roberts L., Gullberg M., Indefrey P.	Online pronoun resolution in L2 discourse: L1 influence and general learner effects	2008	Studies in Second Language Acquisition	30	3	333–357	10.1017/S0272263108080480

#	Authors	Title	Year	Journal	Volume	Issue	Page	DOI
109	O'Rourke B.	The other C in CMC: what alternative data sources can tell us about text-based synchronous computer mediated communication and language learning	2008	Computer Assisted Language Learning	21	3	227–251	10.1080/09588220802090253
110	Kaushanskaya M., Marian V.	Bilingual language processing and interference in bilinguals: evidence from eye tracking and picture naming	2007	Language Learning	57	1	119–163	10.1111/j.1467-9922.2007.00401.x
111	Marian V., Spivey M.	Bilingual and monolingual processing of competing lexical items	2003	Applied Psycholinguistics	24	2	173–193	10.1017/S0142716403000092

Appendix E. Coding Scheme Used to Code the Articles

Category and Variable	Value	Definition and Description	Reference
1. Administrative information			NA
Author	Open	Authors of the study	
Year	Open	The year in which the research was published	
Title	Open	The title of the paper	
Journal	Open	Journal where the study was published	
2. Area of application			Riazi et al. (2018)
Research area	Open	The research area and/or topics of the study	
Research aim	Open	Research aims that were associated with eye-tracking	
Research question	Open	Research questions that were associated with eye-tracking	
3. Cognitive mechanism(s) inferred			Holmqvist et al. (2011); Rayner (2009); Son et al. (2021)
Cognitive mechanism	Open	The type of cognitive mechanism inferred from eye-tracking	
Rationale	Open	Reasons for using eye-tracking to infer this cognitive mechanism	

Category and Variable	Value	Definition and Description	Reference
4. Eye-tracking measure			
Eye-tracking measure type(s)	1. Fixation temporal 2. Fixation count 3. Fixation spatial 4. Saccade temporal 5. Saccade count 6. Saccade spatial 7. Dwell temporal 8. Dwell count 9. Dwell spatial 10. Regression temporal 11. Regression count 12. Regression spatial 13. Skip 14. Pupillometry 15. Blink 16. Gaze patterns	The type of gaze behavior measures used for analysis	Godfroid (2019); Holmqvist et al. (2011); Lai et al. (2013)
Others	Open		
Definition or calculation of each gaze behavior measure	Open	The definition or calculation of the used eye-tracking measure	
5. Study context and participant demographics			Carter and Luke (2020); Fiedler et al. (2020); Godfroid (2019); Holmqvist et al. (2023)
Sample size	1. Below 50 2. 50 to 100 3. Above 100 4. Absent (Not stated in paper)	The number of participants	
Gender distribution	1. Reported 2. Absent (Not stated in paper)	The number of female/male participants	
Age	1. Child (0–12) 2. Teen (13–18) 3. Adult (18+) 4. Multiple 5. Absent (Not stated in paper)	The age of participants	
L1	Open	First language(s) of L2 participants	
Target L2	Open	Target L2(s) of the study	
L2 proficiency	1. No previous knowledge 2. Beginner 3. Intermediate 4. Advanced 5. Mixed 6. Descriptive 7. Absent (Not stated in paper)	L2 proficiency level of L2 participants	
Neurological condition	1. Reported 2. Absent (Not stated in paper)	The neurological condition of the participants	
Visual condition	1. Reported 2. Absent (Not stated in paper)	The vision of the participants	
Hearing condition	1. Reported 2. Absent (Not stated in paper)	The hearing of the participants	
Research site	Open	Country or region where the experiment was conducted (not author's affiliations)	

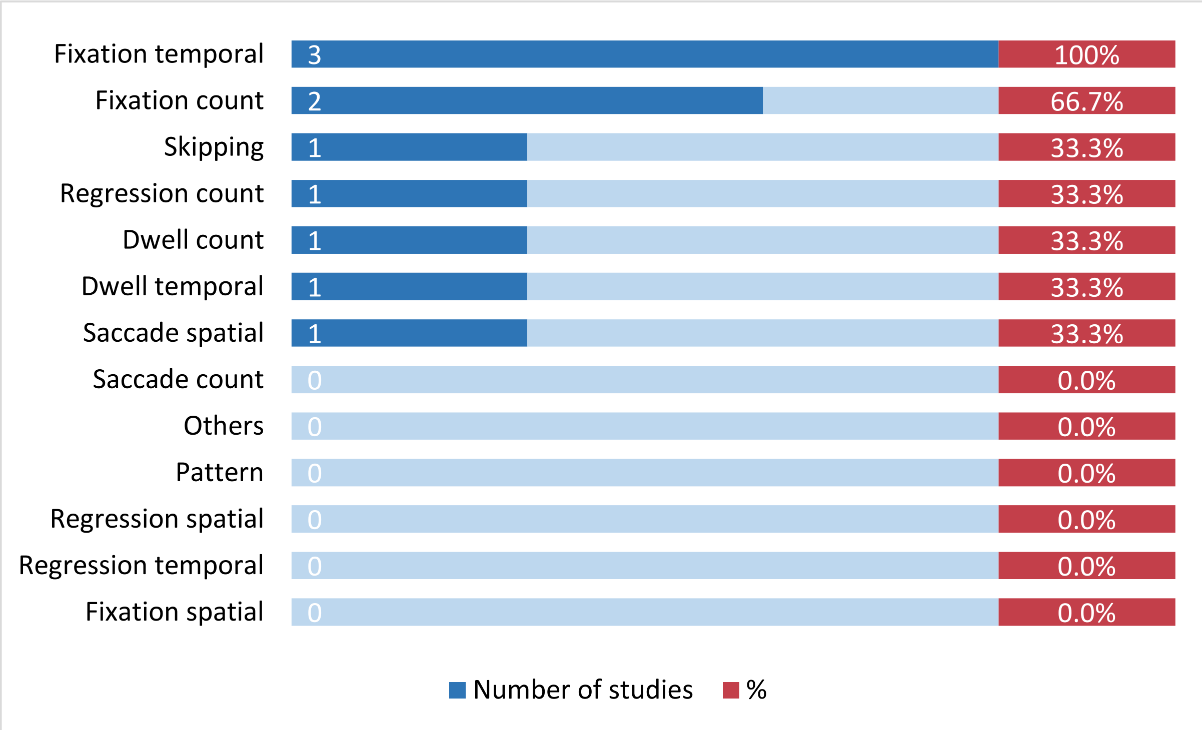
Category and Variable	Value	Definition and Description	Reference
6. Visual stimuli			Conklin and Pellicer-Sánchez (2016); Fiedler et al. (2020); Holmqvist et al. (2023); Spinner et al. (2013)
Visual stimuli type	1. Text 2. Image 3. Text with image 4. Video 5. Video with text 6. Mixed type 7. Absent (Not stated in paper)	The type of materials that participants read/view during the eye-tracking experiment	
Font type	Open	The font type or style of written words	
Font size	1. Reported 2. Absent (Not stated in paper)	The font size of written words	
Text spacing	Open	The text spacing of written words	
Image size	1. Reported 2. Absent (Not stated in paper)	The size of the visuals	
Areas of interests	1. Reported 2. Absent (Not stated in paper)	The content, size, or position of the AOI	
7. Apparatus			Carter and Luke (2020); Fiedler et al. (2020); Holmqvist et al. (2023); King et al. (2019)
Eye-tracking equipment			
Commercial/non-commercial	1. Commercial 2. Non-commercial 3. Absent (Not stated in paper)		
Type	Open	The type of eye-tracking equipment used to collect data	
Brand/manufacturer	Open	The brand or manufacturer of the eye-tracking equipment	
Model	1. Reported 2. Absent (Not stated in paper)	The model of the eye-tracking equipment	
Data sampling frequency	Open	The sampling rate of the eye tracker in Hz	
Number of eyes tracked	1. Monocular 2. Binocular 3. Absent (Not stated in paper)	The number of eyes tracked by the eye tracker	
Head stabilization			
Head movement condition	1. Restrained 2. Unrestrained 3. Absent (Not stated in paper)	The condition of participants' head movement during the eye-tracking experiment	
Monitor			
Display monitor	1. Reported 2. Absent (Not stated in paper)	The brand, pixel resolution, or size of the monitor that was used to present the experimental materials to participants.	
8. Analysis software			Carter and Luke (2020); Fiedler et al. (2020); Holmqvist et al. (2011); Holmqvist et al. (2023); King et al. (2019)
Type of software used	1. Proprietary 2. External vendor (3rd party) 3. Open-source 4. User-developed 5. Absent (Not stated in paper)	The type of the software used to process the raw eye-tracking data	
Name (and version)	Open	Name and version of the software	

Category and Variable	Value	Definition and Description	Reference
9. Eye-tracking data			Carter and Luke (2020); Fiedler et al. (2020); Holmqvist et al. (2011); Holmqvist et al. (2023); King et al. (2019)
Eye data source	1. One eye only 2. Averaged from both eyes 3. Both eyes (i.e., measured independently) 4. Absent (Not stated in paper)	The eye data source used for data analysis	
Data quality	1. Reported 2. Absent (Not stated in paper)	The quality of the eye-tracking data, such as accuracy and track loss	
10. Data pre-processing procedures			Carter and Luke (2020); Fiedler et al. (2020); Godfroid (2019); Holmqvist et al. (2011); Holmqvist et al. (2023); King et al. (2019)
Data interpolation	1. Reported 2. Absent (Not stated in paper)	The interpolation of missing data	
Noise reduction filter	1. Reported 2. Absent (Not stated in paper)	A filter that aims to move all variation in the recorded data that does not derive from true eye movement	
Techniques for parsing eye movements	1. Manual labelling 2. Automatic algorithm 3. Semi-automatic algorithm 4. Absent (Not stated in paper)	The technique used to parse eye movements from the stream of data samples	
Fixation threshold	Open	The minimum duration or dispersion threshold	
Velocity threshold	Open	The velocity threshold is the eye-movement velocity that must be exceeded for a saccade to be detected	

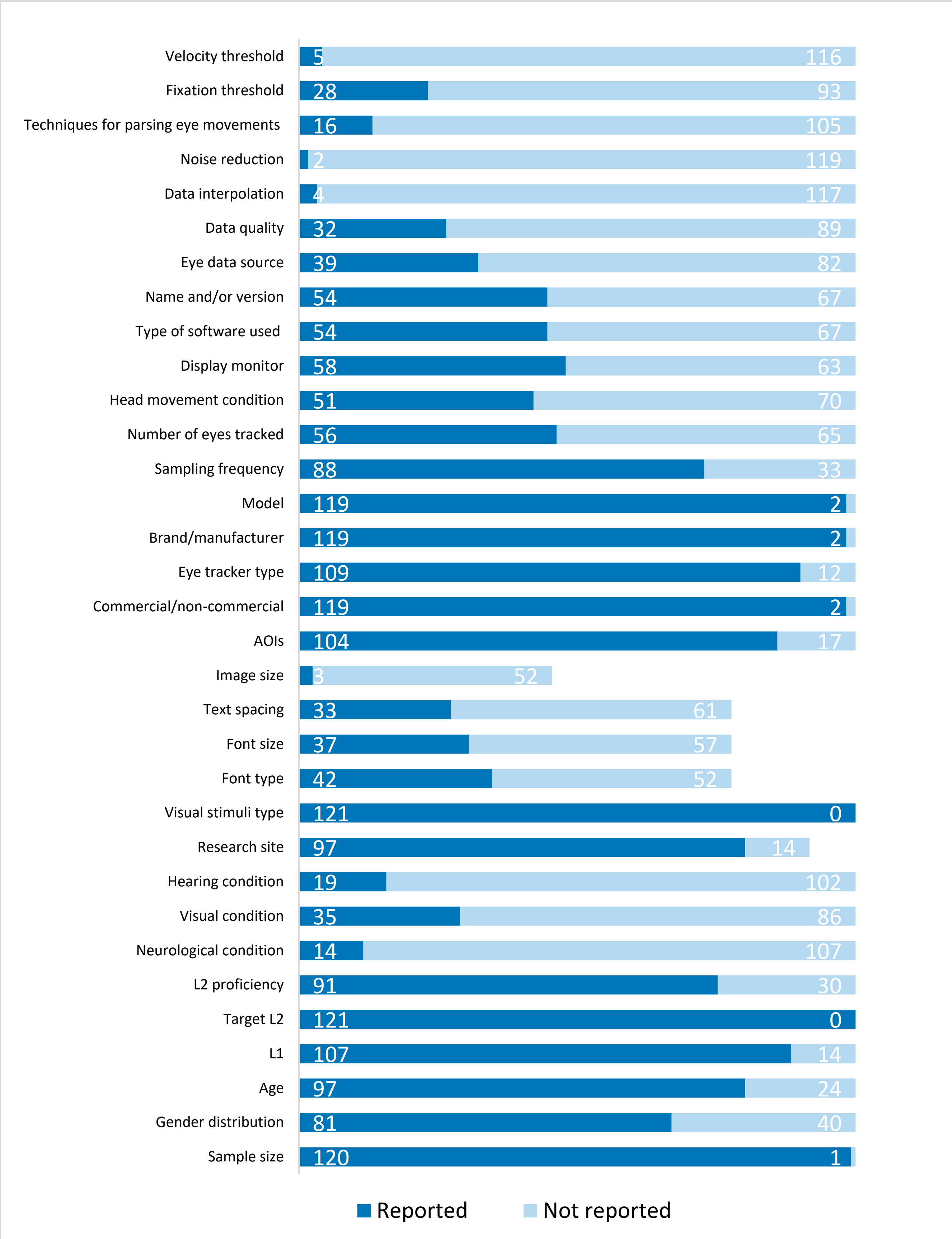
Appendix F. Eye-Tracking Measure Types across Research Areas

Research Area	Subarea		Fixation			Saccade		Dwell		Regression			Skip	Pattern	Others
			Temporal	Count	Spatial	Count	Spatial	Temporal	Count	Temporal	Count	Spatial			
		N	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Grammar	Grammar learning and instruction	17	12 (70.6%)	8 (47.1%)	0	0	0	1 (5.9%)	3 (17.6%)	0	0	0	1 (5.9%)	0	0
	Grammatical processing	10	7 (70.0%)	3 (30.0%)	0	0	0	0	0	5 (50.0%)	3 (30.0%)	0	0	0	0
Vocabulary	Vocabulary learning and instruction	17	17 (100.0%)	9 (52.9%)	0	0	0	0	0	1 (5.9%)	3 (17.6%)	0	1 (5.9%)	0	0
	Bilingual word recognition	5	1 (20.0%)	4 (80.0%)	0	0	0	0	0	1 (20.0%)	0	0	1 (20.0%)	0	0
	Formulaic language processing	2	2 (100.0%)	1 (50.0%)	0	0	0	0	0	1 (50.0%)	0	0	1 (50.0%)	0	0
	Conceptual transfer	1	1 (100.0%)	0	0	0	0	0	1 (100.0%)	0	0	0	0	0	0
	Strategy use	1	0	0	0	0	0	0	0	0	0	0	0	0	1 (100.0%)
Reading	Multimodal reading	7	7 (100.0%)	5 (71.4%)	0	1 (14.3%)	0	3 (42.9%)	0	0	0	0	2 (28.6%)	0	0
	Reading behavior	7	6 (85.7%)	3 (42.9%)	1 (14.3%)	0	1 (14.3%)	2 (28.6%)	0	0	0	0	1 (14.3%)	2 (28.6%)	1 (14.3%)
	Reading test	3	3 (100.0%)	2 (66.7%)	0	0	0	2 (66.7%)	3 (100.0%)	0	0	0	0	1 (33.3%)	1 (33.3%)
Listening	Predictive language processing	9	2 (22.2%)	7 (77.8%)	0	0	0	0	0	0	0	0	0	0	0
	Listening test	4	1 (25.0%)	2 (50.0%)	0	0	0	4 (100.0%)	2 (50.0%)	0	0	0	0	0	0
	Prosody	3	1 (33.3%)	2 (66.7%)	0	0	0	0	0	0	0	0	0	0	0
Writing		12	4 (33.3%)	2 (16.7%)	1 (8.3%)	0	0	0	1 (8.3%)	0	1 (8.3%)	2 (16.7%)	0	2 (16.7%)	3 (25.0%)
Validity		6	3 (50.0%)	2 (33.3%)	0	0	1 (16.7%)	2 (33.3%)	0	0	2 (33.3%)	0	1 (16.7%)	1 (16.7%)	0
Speaking		2	1 (50.0%)	1 (50.0%)	0	0	0	0	1 (50.0%)	0	0	0	0	0	0
Phonology		2	0	0	0	0	0	2 (100.0%)	0	0	0	0	0	0	0
Mixed areas		3	3 (100.0%)	3 (100.0%)	0	1 (33.3%)	1 (33.3%)	1 (33.3%)	0	0	2 (66.7%)	1 (33.3%)	0	0	0
Total		111	71 (64.0%)	54 (48.6%)	2 (1.8%)	2 (1.8%)	3 (2.7%)	17 (15.3%)	11 (9.9%)	8 (7.2%)	11 (9.9%)	3 (2.7%)	8 (7.2%)	6 (5.4%)	6 (5.4%)

Appendix G. The Type of Eye-Tracking Measure Used to Infer Cognitive Load



Appendix H. Reporting Practices in L2 Eye-Tracking Studies Reviewed



Appendix I. Details of Reporting Practices

Table A1. Sample Size, Gender Distribution, and Age of the Sample (*N* = 121).

Variable		*N*	%
Sample size	P (0–50)	64	52.9%
	P (51–100)	46	38.0%
	P (100+)	10	8.3%
	Not reported	1	0.8%
Gender distribution	Reported	81	66.9%
	Not reported	40	33.1%
Age	Child (0–12)	6	5.0%
	Teens (13–17)	2	1.7%
	Adult (18+)	80	66.1%
	Mixed	9	7.4%
	Not reported	24	19.8%

Table A2. First Language, Target L2, and L2 Proficiency of the Sample (*N* = 121).

Variable		*N*	%
First language	English	19	15.7%
	Korean	10	8.3%
	Dutch	8	6.6%
	Chinese (Mandarin)	7	5.8%
	German	5	4.1%
	Japanese	5	4.1%
	Russian	3	2.5%
	Spanish	3	2.5%
	Arabic	2	1.7%
	Greek	2	1.7%
	Sinhala	2	1.7%
	Behasa Malayu	1	0.8%
	Catalan	1	0.8%
	Finnish	1	0.8%
	French	1	0.8%
	Italian Sign language	1	0.8%
	Swedish	1	0.8%
	Turkish	1	0.8%
	More than one L1	34	28.1%
	Not reported	14	11.6%
Target L2	English	82	67.8%
	Spanish	9	7.4%
	French	8	6.6%
	Artificial language	4	3.3%
	German	4	3.3%
	Italian	3	2.5%
	Dutch	2	1.7%
	Latin	2	1.7%
	American Sign Language	1	0.8%
	Chinese	1	0.8%
	Finnish	1	0.8%
	Japanese	1	0.8%
	Palenquero	1	0.8%
	More than one L2	2	1.7%
	Not reported	0	0.0%

Table A2. *Cont.*

Variable		*N*	%
L2 proficiency	No previous knowledge	8	6.6%
	Beginner	3	2.5%
	Intermediate	10	8.3%
	Advanced	8	6.6%
	Mixed levels	21	17.4%
	Descriptive	41	33.9%
	Not reported	30	24.8%

Table A3. Neurological and Physical Condition Relevant to L2 Eye-Tracking Studies (*N* = 121).

Variable		*N*	%
Neurological condition	Reported	14	11.6%
	Not reported	107	88.4%
Visual condition	Reported	35	28.9%
	Not reported	86	71.1%
Hearing condition	Reported	19	15.7%
	Not reported	102	84.3%

Table A4. Research Site Where the Study Was Conducted (*N* = 111).

Research Site	*N*	%
United States	30	27.0%
United Kingdom	16	14.4%
Japan	6	5.4%
Belgium	5	4.5%
Korea	5	4.5%
Spain	3	2.7%
Canada	2	1.8%
Dutch	2	1.8%
Germany	2	1.8%
Netherlands	2	1.8%
New Zealand	2	1.8%
Sri Lanka	2	1.8%
Taiwan	2	1.8%
Australia	1	0.9%
China	1	0.9%
Columbia	1	0.9%
Finland	1	0.9%
Italy	1	0.9%
Malaysia	1	0.9%
Turkey	1	0.9%
More than one research site	11	9.9%
Not reported	14	12.6%

Table A5. Description of the Visual Stimuli and the AOI.

Variable		*N*	%
Stimuli type	Text	66	55.5%
	Image	21	17.6%
	Text with image	12	10.1%
	Video	6	5.0%
	Video with text	7	5.9%
	More than one type	9	7.6%
	Not reported	0	0.0%
Text stimuli (*N* = 94)			
Font type	Courier New	12	12.9%
	Arial	6	6.5%
	Calibri	6	6.5%
	Consolas	6	6.5%
	Times New Roman	6	6.5%
	Courier	2	2.2%
	Verdana	2	2.2%
	Monospace	1	1.1%
	Monotype font	1	1.1%
	Not reported	52	55.9%
Font size	Reported	37	39.4%
	Not reported	57	60.6%
Text spacing	Single word/line	16	17.0%
	1.5-line spacing	2	2.1%
	Double spaced	9	9.6%
	Triple spaced	1	1.1%
	Others	5	5.3%
	Not reported	61	64.9%
Image stimuli (*N* = 55)			
Size	Reported	3	5.5%
	Absent	52	94.5%

Table A6. Description of the Eye-Tracking Equipment (*N* = 121).

Variable		*N*	%
Commercial/non-commercial	Commercial	118	97.5%
	Non-commercial	1	0.8%
	Not reported	2	1.7%
Eye tracker type	Screen-based	64	52.9%
	Head mounted	17	14.0%
	Desk/desktop/tabletop mounted	14	11.6%
	Remote	8	6.6%
	Tower mounted	4	3.3%
	Mobile/portable	2	1.7%
	Not reported	12	9.9%
Brand/manufacturer	SR Research	60	49.6%
	Tobii	42	34.7%
	SMI	5	4.1%
	GazePoint	4	3.3%
	ISCAN	3	2.5%
	EyeTech	2	1.7%
	EyeNTNU	1	0.8%
	FaceLAB	1	0.8%
	LC Technologies	1	0.8%
	Pupil Dev	1	0.8%
	Not reported	2	1.7%
Model	Reported	119	98.3%
	Not reported	2	1.7%

Table A6. *Cont.*

Variable		*N*	%
Sampling frequency	1000 Hz	24	19.8%
	60 Hz	18	14.9%
	120 Hz	14	11.6%
	500 Hz	14	11.6%
	50 Hz	5	4.1%
	300 Hz	5	4.1%
	250 Hz	3	2.5%
	30 Hz	2	1.7%
	150 Hz	1	0.8%
	180 Hz	1	0.8%
	Others	2	1.7%
	Not reported	33	27.3%
Number of eyes tracked	Monocular	39	32.2%
	Binocular	17	14.0%
	Not reported	65	53.7%

Table A7. Description of Head Movement and the Monitor (*N* = 121).

Variable		*N*	%
Head movement	Restrained	33	27.3%
	Unrestrained	18	15.7%
	Not reported	70	57.0%
The monitor	Reported	58	47.9%
	Not reported	63	52.1%

Table A8. Description of the Software Used to Preprocess the Raw Eye-Tracking Data (*N* = 121).

Variable		*N*	%
Type of software used	Proprietary	48	39.7%
	Open source	3	2.5%
	User-developed	2	1.7%
	External vender (3rd party)	1	0.8%
	Not reported	67	55.4%
Name and/or version	Reported	54	44.6%
	Not reported	67	55.4%

Table A9. Description of the Eye-tracking Data: Data Source and Data Quality.

Variable		*N*	%
Eye data source	One eye only	38	31.4%
	Averaged from both eyes	1	0.8%
	Not reported	82	67.8%
Data quality	Reported	32	26.4%
	Not reported	89	73.6%

Table A10. Description of Data Pre-processing Procedures (*N* = 121).

Variable		*N*	%
Data interpolation	Reported	4	3.3%
	Not reported	117	96.7%
Noise reduction	Reported	2	1.7%
	Not reported	119	98.3%
Techniques for parsing eye movements	Automatic algorithm	13	10.7%
	Manual labelling	2	1.7%
	Semi-automatic algorithm	1	0.8%
	Not reported	105	86.8%
Fixation threshold	50 ms	1	0.8%
	60 ms	5	4.1%
	70 ms	1	0.8%
	80 ms	10	8.3%
	100 ms	7	5.8%
	120 ms	1	0.8%
	140 ms	1	0.8%
	150 ms	1	0.8%
	Not reported	94	77.7%
Velocity threshold	30°/s	4	3.3%
	Default setting	1	0.8%
	Not reported	116	95.9%

References

Abdel Latif, Muhammad M. M. 2019. Eye-tracking in recent L2 learner process research: A review of areas, issues, and methodological approaches. *System* 83: 25–35. [CrossRef]

Alemdag, Ecenaz, and Kursat Cagiltay. 2018. A systematic review of eye tracking research on multimedia learning. *Computers and Education* 125: 413–28. [CrossRef]

Alhazmi, Khaled, James Milton, and Stephen Johnston. 2019. Examining 'vowel blindness' among native Arabic speakers reading English words from the perspective of eye-tracking. *System* 80: 235–45. [CrossRef]

Aryadoust, Vahid. 2020. Dynamics of item reading and answer changing in two hearings in a computerized while-listening performance test: An eye-tracking study. *Computer Assisted Language Learning* 33: 510–37. [CrossRef]

Aryadoust, Vahid, and Bee Hoon Ang. 2021. Exploring the frontiers of eye tracking research in language studies: A novel co-citation Scientometric review. *Computer Assisted Language Learning* 34: 898–933. [CrossRef]

Aryadoust, Vahid, and Stacy Foo. 2023. An eye-tracking investigation of visual search strategies and test performance of L1 and L2 listening test takers. *Research and Practice in Technology Enhanced Learning* 18: 9. [CrossRef]

Aryadoust, Vahid, Li Ying Ng, and Hiroki Sayama. 2021. A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing* 38: 6–40. [CrossRef]

Aryadoust, Vahid, Stacy Foo, and Li Ying Ng. 2022. What can gaze behaviors, neuroimaging data, and test scores tell us about test method effects and cognitive load in listening assessments? *Language Testing* 39: 56–89. [CrossRef]

Baker, Monya. 2016. Reproducibility crisis. *Nature* 533: 353–66. [CrossRef]

Batty, Aaron Olaf. 2021. An eye-tracking study of attention to visual cues in L2 listening tests. *Language Testing* 38: 511–35. [CrossRef]

Bax, Stephen. 2013. The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing* 30: 441–65. [CrossRef]

Bax, Stephen. 2015. Using eye-tracking to research the cognitive processes of multinational readers during an IELTS reading test. *IELTS Research Reports Online Series* 2: 1–21. Available online: https://www.ielts.org/-/media/research-reports/ielts_online_rr_2015-2.ashx (accessed on 20 December 2023).

Bax, Stephen, and Sathena Chan. 2019. Using eye-tracking research to investigate language test validity and design. *System* 83: 64–78. [CrossRef]

Beatty, Jackson, and Brennis Lucero-Wagoner. 2000. The pupillary system. In *Handbook of Psychophysiology*. Cambridge: Cambridge University Press, pp. 142–62, ISBN 978-0-521-62634-7.

Carter, Benjamin T., and Steven G. Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155: 49–62. [CrossRef]

Conklin, Kathy, Ana Pellicer-Sánchez, and Gareth Carrol. 2018. *Eye-Tracking: A Guide for Applied Linguistics Research*. Cambridge: Cambridge University Press.

Conklin, Kathy, and Ana Pellicer-Sánchez. 2016. Using eye-tracking in applied linguistics and second language research. *Second Language Research* 32: 453–67. [CrossRef]

Connell, Katrina, Simone Hüls, Maria Teresa Martínez-García, Zhen Qin, Seulgi Shin, Hanbo Yan, and Annie Tremblay. 2018. English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning* 68: 635–68. [CrossRef]
Cooper, Harris, Larry V. Hedges, and Jeffrey C. Valentine. 2019. *The Handbook of Research Synthesis and Meta-Analysis*. Manhattan: Russell Sage Foundation.
Crowther, Dustin, Susie Kim, Jongbong Lee, Jungmin Lim, and Shawn Loewen. 2021. Methodological synthesis of cluster analysis in second language research. *Language Learning* 71: 99–130. [CrossRef]
Cunnings, Ian, Georgia Fotiadou, and Ianthi Tsimpli. 2017. Anaphora resolution and reanalysis during L2 sentence processing: Evidence from the visual world paradigm. *Studies in Second Language Acquisition* 39: 621–52. [CrossRef]
Derrick, Deirdre J. 2016. Instrument reporting practices in second language research. *TESOL Quarterly* 50: 132–53. [CrossRef]
Duchowski, Andrew T. 2017. *Eye Tracking Methodology*. Berlin/Heidelberg: Springer International Publishing, ISBN 978-3-319-57881-1. [CrossRef]
Dussias, Paola E. 2010. Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics* 30: 149–66. [CrossRef]
Eckstein, Maria K., Belén Guerra-Carrillo, Alison T. Miller Singley, and Silvia A. Bunge. 2017. Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience* 25: 69–91. [CrossRef] [PubMed]
Elgort, Irina, Marc Brysbaert, Michaël Stevens, and Eva Van Assche. 2018. Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition* 40: 341–66. [CrossRef]
Esteve-Gibert, Núria, and Carmen Muñoz. 2021. Preschoolers benefit from a clear sound-referent mapping to acquire nonnative phonology. *Applied Psycholinguistics* 42: 77–100. [CrossRef]
Fiedler, Susann, Michael Schulte-Mecklenbeck, Frank Renkewitz, and Jacob Lund Orquin. 2020. Guideline for Reporting Standards of Eye-Tracking Research in Decision Sciences. PsyArXiv. Available online: https://osf.io/preprints/psyarxiv/f6qcy (accessed on 20 December 2023).
Flecken, Monique, Mary Carroll, Katja Weimar, and Christiane Von Stutterheim. 2015. Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French-German L2 users. *The Modern Language Journal* 99: 100–22. [CrossRef]
Fujita, Hiroki, and Ian Cunnings. 2021. Lingering misinterpretation in native and nonnative sentence processing: Evidence from structural priming. *Applied Psycholinguistics* 42: 475–504. [CrossRef]
Godfroid, Aline. 2019. *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide*, 1st ed. London: Routledge, ISBN 978-1-315-77561-6.
Godfroid, Aline, and Bronson Hui. 2020. Five common pitfalls in eye-tracking research. *Second Language Research* 36: 277–305. [CrossRef]
Godfroid, Aline, Frank Boers, and Alex Housen. 2013. An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition* 35: 483–517. [CrossRef]
Henderson, John. 2003. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences* 7: 498–504. [CrossRef]
Hess, E. H., and J. M. Polt. 1964. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science* 143: 1190–92. [CrossRef]
Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press, ISBN 978-0-19-969708-3.
Holmqvist, Kenneth, Saga Lee Örbom, Ignace T. C. Hooge, Diederick C. Niehorster, Robert G. Alexander, Richard Andersson, Jeroen S. Benjamins, Pieter Blignaut, Anne-Marie Brouwer, Lewis L. Chuang, and et al. 2023. Eye tracking: Empirical foundations for a minimal reporting guideline. *Behavior Research Methods* 55: 364–416. [CrossRef]
Huettig, Falk, Joost Rommers, and Antje S. Meyer. 2011. Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* 137: 151–71. [CrossRef]
Hung, Yueh-Nu, Hui-Yu Kuo, and Shih-Chieh Liao. 2020. Seeing what they see: Elementary EFL students reading science texts. *RELC Journal* 51: 397–411. [CrossRef]
Indrarathne, Bimali, and Judit Kormos. 2017. Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition* 39: 401–30. [CrossRef]
Issa, Bernard I., and Kara Morgan-Short. 2019. Effects of external and internal attentional manipulations on second language grammar development: An eye-tracking study. *Studies in Second Language Acquisition* 41: 389–417. [CrossRef]
Just, Marcel A., and Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87: 329–54. [CrossRef] [PubMed]
Keating, Gregory D. 2009. Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning* 59: 503–35. [CrossRef]
Keating, Gregory D., and Jill Jegerski. 2015. Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition* 37: 1–32. [CrossRef]

Kho, Shermaine Qi En, Vahid Aryadoust, and Stacy Foo. 2023. An eye-tracking investigation of the keyword-matching strategy in listening assessment. *Education and Information Technologies* 28: 3739–63. [CrossRef]

Kim, Hyunwoo, and Theres Grüter. 2021. Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition* 43: 133–54. [CrossRef]

King, Andy J., Nadine Bol, R. Glenn Cummins, and Kevin K. John. 2019. Improving visual behavior research in communication science: An overview, review, and reporting recommendations for using eye-tracking methods. *Communication Methods and Measures* 13: 149–77. [CrossRef]

Krejtz, Krzysztof, Andrew T. Duchowski, Anna Niedzielska, Cezary Biele, and Izabela Krejtz. 2018. Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLoS ONE* 13: e0203629. [CrossRef] [PubMed]

Lai, Meng-Lung, Meng-Jung Tsai, Fang-Ying Yang, Chung-Yuan Hsu, Tzu-Chien Liu, Silvia Wen-Yu Lee, Min-Hsien Lee, Guo-Li Chiou, Jyh-Chong Liang, and Chin-Chung Tsai. 2013. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review* 10: 90–115. [CrossRef]

Lee, Minjin, and Andrea Révész. 2020. Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition* 42: 625–51. [CrossRef]

Lee, Minjin, and Jookyoung Jung. 2021. Effects of textual enhancement and task manipulation on L2 learners' attentional processes and grammatical knowledge development: A mixed methods study. *Language Teaching Research* 136216882110346. [CrossRef]

Lee, Shinhye, and Paula Winke. 2018. Young learners' response processes when taking computerized tasks for speaking assessment. *Language Testing* 35: 239–69. [CrossRef]

Liberati, Alessandro, Douglas G. Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C. Gøtzsche, John P. A. Ioannidis, Mike Clarke, P. J. Devereaux, Jos Kleijnen, and David Moher. 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine* 6: e1000100. [CrossRef]

Low, Andralyn Rui Lin, and Vahid Aryadoust. 2023. Investigating test-taking strategies in listening assessment: A comparative study of eye-tracking and self-report questionnaires. *International Journal of Listening* 37: 93–112. [CrossRef]

Makel, Matthew C., and Jonathan A. Plucker. 2014. Facts are more important than novelty: Replication in the education sciences. *Educational Researcher* 43: 304–16. [CrossRef]

Marian, Viorica, and Michael Spivey. 2003. Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics* 24: 173–93. [CrossRef]

Marsden, Emma, and Kara Morgan-Short. 2023. (Why) Are open research practices the future for the study of language learning? *Language Learning* 73: 344–87. [CrossRef]

Marsden, Emma, Kara Morgan-Short, Sophie Thompson, and David Abugaber. 2018. Replication in second language research: Narrative and systematic reviews and recommendations for the field: Replication in second language research. *Language Learning* 68: 321–91. [CrossRef]

Meghanathan, Radha Nila, Cees van Leeuwen, and Andrey R. Nikolaev. 2015. Fixation duration surpasses pupil size as a measure of memory load in free viewing. *Frontiers in Human Neuroscience* 8: 1063. [CrossRef] [PubMed]

Mitsugi, Sanako. 2017. Incremental comprehension of Japanese passives: Evidence from the visual-world paradigm. *Applied Psycholinguistics* 38: 953–83. [CrossRef]

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and PRISMA Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine* 151: 264–69. [CrossRef]

Montero Perez, Maribel. 2019. Pre-learning vocabulary before viewing captioned video: An eye-tracking study. *The Language Learning Journal* 47: 460–78. [CrossRef]

Montero Perez, Maribel, Elke Peters, and Piet Desmet. 2015. Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal* 99: 308–28. [CrossRef]

Nahatame, Shingo. 2021. Text readability and processing effort in second language reading: A computational and eye-tracking investigation. *Language Learning* 71: 1004–43. [CrossRef]

Norris, John M., Luke Plonsky, Steven J. Ross, and Rob Schoonen. 2015. Guidelines for reporting quantitative methods and results in primary research. *Language Learning* 65: 470–76. [CrossRef]

Orquin, Jacob L., and Kenneth Holmqvist. 2018. Threats to the validity of eye-movement research in psychology. *Behavior Research Methods* 50: 1645–56. [CrossRef]

Paas, Fred, Juhani E. Tuovinen, Huib Tabbers, and Pascal W. M. Van Gerven. 2003. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist* 38: 63–71. [CrossRef]

Pellicer-Sánchez, Ana, Elsa Tragant, Kathy Conklin, Michael Rodgers, Raquel Serrano, and Ángels Llanes. 2020. Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition* 42: 577–98. [CrossRef]

Pellicer-Sánchez, Ana, Kathy Conklin, and Laura Vilkaitė-Lozdienė. 2021a. The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning* 71: 162–203. [CrossRef]

Pellicer-Sánchez, Ana, Kathy Conklin, Michael P. H. Rodgers, and Fabio Parente. 2021b. The effect of auditory input on multimodal reading comprehension: An examination of adult readers' eye movements. *The Modern Language Journal* 105: 936–56. [CrossRef]

Perkhofer, Lisa, and Othmar Lehner. 2019. Using gaze behavior to measure cognitive load. In *Information Systems and Neuroscience*. Edited by F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger and A. B. Randolph. Lecture Notes in Information Systems and Organisation. Berlin/Heidelberg: Springer International Publishing, pp. 73–83, ISBN 978-3-030-01086-7. [CrossRef]

Petticrew, Mark, and Helen Roberts. 2008. *Systematic Reviews in the Social Sciences: A Practical Guide*. Hoboken: John Wiley and Sons.

Plonsky, Luke, and Talip Gonulal. 2015. Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning* 65: 9–36. [CrossRef]

Rahal, Rima-Maria, and Susann Fiedler. 2019. Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology* 85: 103842. [CrossRef]

Ranalli, Jim, Hui-Hsien Feng, and Evgeny Chukharev-Hudilainen. 2019. The affordances of process-tracing technologies for supporting L2 writing instruction. *Language Learning and Technology* 23: 1–11.

Rayner, Keith. 2009. The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology* 62: 1457–506. [CrossRef]

Rayner, Keith, Kathryn H. Chace, Timothy J. Slattery, and Jane Ashby. 2006. Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading* 10: 241–55. [CrossRef]

Révész, Andrea, Marije Michel, and Minjin Lee. 2019. Exploring second language writers' pausing and revision behaviors: A mixed-methods study. *Studies in Second Language Acquisition* 41: 605–31. [CrossRef]

Révész, Andrea, Rebecca Sachs, and Mika Hama. 2014. The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts: Task complexity, input frequency, and recasts. *Language Learning* 64: 615–50. [CrossRef]

Riazi, A. Mehdi, Hessameddin Ghanbar, and Ismaeil Fazel. 2020. The contexts, theoretical and methodological orientation of EAP research: Evidence from empirical articles published in the Journal of English for Academic Purposes. *Journal of English for Academic Purposes* 48: 100925. [CrossRef]

Riazi, Mehdi, Ling Shi, and John Haggerty. 2018. Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing* 41: 41–54. [CrossRef]

Roberts, Leah, and Anna Siyanova-Chanturia. 2013. Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition* 35: 213–35. [CrossRef]

Roberts, Leah, Marianne Gullberg, and Peter Indefrey. 2008. Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition* 30: 333–57. [CrossRef]

Schmidtke, Jens. 2018. Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition* 40: 529–49. [CrossRef]

Schotten, Michiel, Wim J. N. Meester, Susanne Steiginga, and Cameron A. Ross. 2017. A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In *Research Analytics*, 1st ed. Edited by F. J. Cantú-Ortiz. Boca Raton: Auerbach Publications, pp. 31–58.

Son, Myeongeun, Jongbong Lee, and Aline Godfroid. 2021. Attention to form and meaning revisited: Insights from eye tracking. *Studies in Second Language Acquisition* 44: 788–817. [CrossRef]

Spinner, Patti, Susan M. Gass, and Jennifer Behney. 2013. Ecological validity in eye-tracking: An empirical study. *Studies in Second Language Acquisition* 35: 389–415. [CrossRef]

Stone, Adam, Laura-Ann Petitto, and Rain Bosworth. 2018. Visual sonority modulates infants' attraction to sign language. *Language Learning and Development* 14: 130–48. [CrossRef]

Strohmaier, Anselm R., Kelsey J. MacKay, Andreas Obersteiner, and Kristina M. Reiss. 2020. Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics* 104: 147–200. [CrossRef]

Suvorov, Ruslan. 2015. The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing* 32: 463–83. [CrossRef]

Traxler, Matthew J., Timothy Banh, Madeline M. Craft, Kurt Winsler, Trevor A. Brothers, Liv J. Hoversten, Pilar Piñar, and David P. Corina. 2021. Word skipping in deaf and hearing bilinguals: Cognitive control over eye movements remains with increased perceptual span. *Applied Psycholinguistics* 42: 601–30. [CrossRef]

van der Wel, Pauline, and Henk van Steenbergen. 2018. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review* 25: 2005–15. [CrossRef]

Wang, Yingxu, Ying Wang, S. Patel, and D. Patel. 2006. A layered reference model of the brain (LRMB). *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 36: 124–33. [CrossRef]

Warren, Paul, Frank Boers, Gina Grimshaw, and Anna Siyanova-Chanturia. 2018. The effect of gloss type on learners' intake of new words during reading: Evidence from eye-tracking. *Studies in Second Language Acquisition* 40: 883–906. [CrossRef]

Wiener, Seth, Kiwako Ito, and Shari R. Speer. 2021. Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinations: An eye-tracking study. *Studies in Second Language Acquisition* 43: 155–80. [CrossRef]

Winke, Paula M. 2013. The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition* 35: 323–52. [CrossRef]

Wong, Wynne, and Kiwako Ito. 2018. The effects of processing instruction and traditional instruction an L2 online processing of the causative construction in French: An eye-tracking study. *Studies in Second Language Acquisition* 40: 241–68. [CrossRef]
Zhou, Wei, Wanwen Ye, and Ming Yan. 2020. Alternating-color words facilitate reading and eye movements among second-language learners of Chinese. *Applied Psycholinguistics* 41: 685–99. [CrossRef]