Coming from different directions: A comparison of the eye movements of English L1 and Arabic L1 speakers reading in English and the implementation of an intensive reading intervention programme

Joan Oakley

B.A. Memorial University of Newfoundland

M.A. Concordia University

This thesis is submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Applied Linguistics

Department of Linguistics and English Language

Lancaster University

February 2018

Acknowledgements

I owe a debt of gratitude to many people for the assistance I have been given over the past six years. First and foremost, I would like to thank my supervisor, Dr. Judit Kormos, who has been a constant source of inspiration, guidance and support from the early stages of course work to the final version of this thesis. I am extremely fortunate to have had such an exemplary mentor.

At the college of the North Atlantic, I would like to thank the former Head of the Office of Applied Research, Dr. Mike Long, who believed in my research from the outset and provided the technology which allowed me to conduct the eye tracking studies. I would also like to thank my former Dean, Jason Rolls, who gave permission to carry out Studies One and Two and was supportive in terms of granting leave to attend the residentials in Lancaster, the eye tracking courses and the viva. I would also like to acknowledge the 36 colleagues who volunteered to be the skilled L1 participants in Study One and the 39 students who participated in the studies. I give a special thanks to Meredith Bishop, Tom McCaughy, Randall Wolansky and Marcin Wawrzyniak who taught the experimental and control groups in Study Two. I would like to thank Shelagh Lariviere who has been a constant source of encouragement and reassurance, Wayne Chisholm, Rodney Bruce, Janice Connolly for ongoing encouragement and David Lukin and John Allen for technical support.

At Lancaster, I would like to thank Dr. Jane Sunderland, the former Director of Studies, my first contact in the PhD Programme who was instrumental in ensuring the cohesiveness of Cohorts 10, 11 and 12, Dr. Tineke Brunfaut who helped me overcome my fear of testing and evaluation and who gave valuable comments on my confirmation panel, Dr. Gareth McCray who was of invaluable assistance in analysing the eye tracking data and Elaine Heron who was always a great source of information. I would like to thank my classmates from Cohorts 11 and 12

especially Jean Jimenez and Attila Wind; your friendship has been instrumental in the successful completion of this thesis.

At Consort World in Dubai, I would like to thank Camille Matta who first showed me how to use the Tobii TX 300 and to Tomas Nilsson who was always there for technical support and information.

At Lund University, I would like to thank Dr. Kenneth Holmqvist and his eye tracking team for the training sessions which were vital in my understanding of eye movement research.

At Tobii Pro, I would like to thank Ricardo Matos, Head of Training for the very informative workshops on the Tobii TX300 held in Frankfurt.

From Concordia University, I would like to thank my two mentors Dr. Jack Upshur, who instilled in me a love of research and Dr. Patsy Lightbown who continues to be a role model.

Finally, I would like to thank my son William who always believed that his mother was not too old to start a PhD.

Abstract

Eye movements have been widely used to investigate cognitive processes during first language (L1) reading and in the last few years it has become more common for applied linguists to use eye tracking technology to examine topics that had previously been investigated using off-line measures (Conklin & Pellicer-Sánchez, 2016). However, few eye movement studies have investigated Arabic L1 speakers reading in English, although there have been numerous studies which showed that this population experiences great difficulty learning to read in English (see, for example, Abu-Rabia, 1997b; Fender, 2003, 2008; Hayes-Harb, 2006;; Randall, 2007; Randall & Groom, 2009; Randall & Meara, 1988; Ryan, 1997; Ryan & Meara, 1991; Saigh & Schmitt, 2012; Thompson-Panos & Thomas-Ruzic, 1983). There is general consensus in the literature that these problems are related to inefficient processing of English vowels. The overall aim of this thesis was to investigate the causes of the reading difficulties I observed with my own students and then to devise a barrage of pedagogical interventions which would remediate these problems. The research was comprised of two studies.

The aim of Study One of the present research was to investigate the differences in eye movements between skilled English L1 (N=36) and Arabic L1 EFL participants (N=39) reading English sentences. Study One found that the Arabic L1 participants displayed eye movements which were significantly different from the patterns exhibited by the English L1 participants, and were indicative of potentially less efficient cognitive processes. These differences were demonstrated in 10 of the 11 metrics calculated. Specifically, the Arabic L1 EFL participants exhibited significantly more and longer fixations than the English L1 participants. They also made significantly more and shorter (forward) saccades than the English L1 group. Furthermore, data

analysis showed a highly significant difference between the two groups in visits on vowels and consonants.

Study One constitutes an important contribution to the literature on the difficulties experienced by Arabic L1 students learning to read in English. It demonstrates that their eye movements are significantly different from those of skilled English L1 readers. Little or no work exists which investigates any differences in allotment of visual attention when comparing the eye movements on vowels and consonants of Arabic L1 and English L1 speakers as they read sentences in English. The finding that the Arabic L1 EFL participants spent more time attending to vowels than did the English L1 participants questions the 'vowel blindness' hypothesis as proposed by Ryan and Meara (1991). This refers to the assumption that Arabic L1 speakers "lack an awareness of the function which vowels perform in English" (Ryan, 1997, p. 189) and consequently do not recognize or attend to them.

The aim of Study Two was to investigate the effects of focused reading interventions on the eye movement patterns and overall reading proficiency of Arabic L1 EFL students. Study Two was a quasi-experimental study which compared two groups of proficiency-matched Arabic L1 EFL learners (N= 39), before and after an intensive reading intervention programme during a 14-week semester at a technical college in Qatar. It included two intact classes in the experimental group (N=20) which received reading interventions consisting of textual enhancement, phonemic awareness, spelling, tracking exercises, rapid word recognition and oral text fluency and two intact classes in the control group which received regular classroom instruction (N=19). The effect of the treatment on reading test scores was analysed using a 2-way repeated ANOVA. Analysis of total reading scores showed a significant main effect for time, but no significant main effect for experimental condition. To investigate the eye movements of the two groups before and after

treatment, this phase of the study used the same eye tracking metrics employed in Study One. Results showed there was no statistically significant interaction between the experimental *group and time*, indicating that both the treatment and control groups showed improvement in their eye tracking measures during the 14 weeks. Study Two is the first study to investigate a barrage of pedagogical interventions on the eye movements and reading proficiency of Arabic L1 EFL students. Although the interventions did not produce statistically significant results, the study provides a building block for future studies using focused pedagogical interventions with this particular group of learners.

Contents

List of tables

List of figures

List of abbreviations

AOI	Areas of Interest
BUiD	British University in Dubai
BYOD	Bring Your Own Device
CEFR	Common European Framework of Reference
CELEX	Communitatis Europeae Lex (European Union law database)
EAP	English for Academic Purposes
ESL	English as a Second Language
EFL	English as a Foreign Language
FCAT	Florida Comprehensive Assessment Test
GCC	Gulf Cooperative Countries
L1	First Language
L2	Second Language
MGR	Mental Graphemic Representations
MPSA	Morphological Processing Spelling Approach
MSA	Modern Standard Arabic
NCTQ	National Council Teacher Quality
NGSL	New General Service List

NH	Noticing Hypothesis
NRP	National Reading Panel
PA	Phonological Awareness
PVL	Preferred viewing location
RAs	Roman Alphabets
RQ	Research Question
RWR	Rapid Word Naming
SLA	Second Language Acquisition
SVR	Simple View of Reading
TAP	Think Aloud Protocol
TE	Textual Enhancement
WRT	Word Recognition Test

Chapter 1: Introduction

1.1 Motivation for the research

Although Second Language Acquisition (SLA) research and language teaching are connected, the relationship is often contentious. While some researchers such as Larsen-Freeman (1998), Larsen-Freeman and Long (1991) argued that one of the key purposes of SLA research is to improve second language teaching, others such as Block (2000) and Crookes (1993) voiced concerns about the extent to which SLA research has actually influenced language teaching. Block (op.cit.) maintained that much of the research which has been conducted in the field of SLA cannot directly be applied to the teaching and learning that goes on in second and foreign language classrooms. Ellis (1997 b; 2001) stated that, although SLA research has made much progress, there is little that teachers can gain from it in terms of practical ideas and "much of the research is no longer directly concerned with pedagogic issues" (Ellis, 2001, p. 45). Nassaji (2012) suggested that the results of SLA research cannot be used by classroom teachers because, in the vast majority of cases, the research is carried out under controlled experimental conditions and results may therefore not be generalizable to the actual teaching and learning contexts. Furthermore, in many cases, researchers who work in SLA research may not be familiar with the day-to-day issues surrounding the classroom and therefore may not investigate areas which teachers consider meaningful. A final reason suggested by Nassaji (op.cit.) for the 'disconnect' between SLA research and the classroom is the use of metalanguage used by researchers which can often be difficult for many teachers to understand. This lack of coordination between researchers and classroom teachers led to what Clark (1994) called the 'theory practice dysfunction'.

This is not so say that studies conducted by researchers are not relevant for SLA pedagogy. There have been numerous studies carried out by researchers which have contributed to the knowledge base of teachers. However, one type of research which can be considered particularly relevant to pedagogical issues is research which is conducted by teachers themselves, provided they are equipped with the means, skills and time to conduct such research. This type of research has been variously termed teacher research, action research, practitioner research, and Exploratory Practice (EP). EP is a form of practitioner research in language education which aims to integrate research, learning and teaching (Hanks, 2015). Developed by Allwright (1993), EP promotes the notion of teachers contemplating their language teaching experiences using typical pedagogic practices as investigative tools (Allwright, 2003; 2005). By incorporating research into pedagogy, EP attempts to address the issue of the demands that research imposes on teachers which might add to their workload. Nassaji (2012) suggested that teachers should conduct research in their own classrooms because by doing so, they can evaluate the relevance of pedagogical issues proposed by SLA, test hypotheses about instructional practices and contribute to theories of instructed SLA. He added that the outcomes of this type of research are more likely to be used by teachers because such an approach examines the practical pedagogic issues that teachers consider important, as opposed to theoretical issues chosen by SLA researchers (Ellis, 1997 a). However, as Block (2000) pointed out, the "entire enterprise is strong in theory but very difficult to carry out in practice" (p. 138). Firstly, in most teaching contexts, teachers are not compensated for the extra work that carrying out research might entail. Additionally, those teachers who do decide to engage in research will invariably find that their results will never have as much impact or influence on the field of teaching as those formulated by professional researchers (Block, op.cit.). I believe I have overcome the last impediment as I am in the enviable position of being both an experienced

classroom teacher and an emergent researcher. Currently, I teach at a technical college in Qatar where for the past 11 years, I have taught EFL students of varying levels of proficiency. No matter what the students' level is, I am consistently faced with learners who are unsuccessful in effectively deciphering the English words on the page, despite the many different methods and techniques I have implemented to help remediate this problem. The students could effectively do scanning exercises where they had to find a predetermined word, phrase or date. They were less effective when search reading which required them to look for words in the same semantic field but not an exact match. Reading carefully to obtain meaning from a text proved much more challenging as was constantly evidenced by reading tests and exams. Oral or written summaries of even the shortest texts provided evidence that my students did not understand what they had read. When reading words aloud, they would pronounce only parts. It was as if they were merely dipping into the word at various places but not reading the word in its entirety. My own anecdotal evidence was confirmed by other EFL teachers. Adding more credence to our classroom observations was empirical evidence provided by numerous researchers in the field. The reading difficulties of Arabic L1 learning to read in English have been widely documented in the literature (see for example, Abu-Rabia, 1997b; Fender, 2003, 2008; Hayes-Harb, 2006; Randall & Meara, 1988; Randall, 2007; Randall & Groom, 2009; Ryan, 1997; Ryan & Meara, 1991; Saigh & Schmitt; 2012; Thompson-Panos & Thomas-Ruzic, 1983). The PhD programme at Lancaster presented me with the impetus to read more extensively about the learning, teaching and assessment of reading, and provided me with the skills I needed to do my own research on what makes reading so challenging for these students.

The impetus for the research reported in this thesis therefore, stems from a practical classroom issue: the literacy-related difficulties experienced by my Arabic L1 (first language) EFL

students. In over 30 years of teaching I had never encountered an L1 population who experienced such problems learning to read and write in English until I moved to the Middle East. I wanted to understand the reason behind my students' difficulties and to find solutions which would hopefully remediate the problems. I wanted to better understand what they were doing differently from students of other L1 backgrounds and from skilled L1 English readers. In addition to exploring studies on cognitive processes in reading, and reviewing the findings on the difficulties experienced by Arabic L1 EFL learners, I turned to the literature on eye tracking. However, I found that the majority of eye tracking research has been carried out in the 'default' mode of reading, where comprehension is "proceeding without difficulty and the eyes are continuing to move forward along a line of text" (Reichle, Warren & McConnell, 2009, p. 9). Therefore, "much of the research which has been conducted with L1 English participants may not be applicable to second language reading" (Bax, 2013, p. 8). Study One investigated the differences in the eye movements of Arabic L1 participants and skilled English L1 participants reading in English. To my knowledge, there have been no eye movement studies to date which have made this comparison to ascertain how 'close' or how 'far away' the Arabic L1 readers are from English L1 readers reading in English. Additionally, little or no work exists which investigates any differences in allotment of visual attention when comparing the eye movements on vowels and on consonants of Arabic L1 and English L1 speakers as they read English sentences. Study Two was conducted with EFL learners and addressed a practical classroom issue: the reading difficulties experienced by Arabic L1 EFL students. The results of this study will make a concrete contribution in terms of pedagogical interventions which may be used to help overcome the reading difficulties experienced by these particular learners.

1.2. Theoretical background to the study

Cognitive processes in first language (L1) reading

The Simple View of Reading (SVR) (Gough & Tunmer, 1986; Hoover& Gough, 1990) proposes that reading comprehension is the product of two basic components: word reading and language (listening) comprehension. Word reading is the ability to read single words out of context by using either a lexical or non-lexical route. The lexical route involves using visual information to retrieve the meaning of the word while the non-lexical route involves translating the word into an auditory representation and using this representation to retrieve the meaning of the word (Coltheart, 2006; Coltheart et al., 1993; 2001; 2005). Reading and language comprehension work together and are both necessary for successful text comprehension (Gough & Tunmer, 1986). An important component of word recognition is orthographic processing which is based on the formation of visual long-term memory representations of letters, letter patterns and sequences of letters that spatially map the temporal sequence of phonemes within words (Ehri, 1992; 1998; 2005). By repeated association of the word's pronunciation with its visual representation, readers develop not only their word-decoding skills but establish a memory trace for the written form of words.

Taxonomies of reading comprehension abilities often categorize the component skills and processes as higher-or lower-level skills. Oakhill et al. (2015) proposed that there are three strands involved in word reading and five strands involved in language comprehension which, when braided together result in efficient and fluent reading comprehension. The lower order processes and components include 1) letter-sound knowledge 2) accurate word decoding and 3) automaticity in decoding. The higher order processes involve 1) activating word meanings 2)

understanding and linking sentences 3) inference making 4) comprehension monitoring and 5) understanding text structure.

The component processes involved in reading comprehension require that the information from word or text knowledge is available and accessible (Cain, Oakhill & Bryant, 2004). Working memory is a temporary storage and processing system that is necessary for a range of cognitive tasks (Baddeley, 1986; Daneman & Carpenter, 1980). Working memory holds the most recently read content and also information retrieved from long-term memory, facilitating its integration into the currently active text. Slow or inaccurate word reading is proposed to affect comprehension by using up too much processing capacity, so that there is little remaining for text comprehension (Hannon & Daneman, 2001; Perfetti, 1985).

Cognitive processes in second language (L2) reading

Grabe (2009) divided the cognitive processes involved in second language reading into two levels: lower-level and higher-level. The lower-level processes involve letter recognition word recognition, syntactic parsing and proposition encoding. The higher-level processes involve integrating information within a text, activating and utilizing background knowledge information and making inferences (Khalifa & Weir, 2009). What distinguishes the lower-level processes from the higher-level processes is the fact that the lower-level processes can become strongly automatized, while the higher-level processes require attentional resources (Grabe, 2009; Khalifa & Weir, op.cit).

Cross-linguistic approach to reading

There are important differences between first and second language reading. Readers in the L1 have years to build language knowledge before they begin to read while L2 readers must develop linguistic resources at the same time as they develop reading comprehension. Second

language readers are working with two languages and so it is likely that they will engage in different types of metalinguistic processing while reading in the L2 (Grabe, 2009). L2 reading, unlike L1 reading, involves two languages and the dual-language involvement implies continual interactions between those languages as well as constant adjustments in accommodating the conflicting demands each language imposes (Koda, 2007). Cowan and Sarmad (1976) proposed that bilingual children have two sets of processing strategies and that transfer will be facilitated according to the similarity of the two languages involved.

1.3 Aims and design of the studies

This thesis presents the findings of two studies. Study One compared the eye movements of skilled English L1 readers and Arabic L1 EFL participants. Eye tracking is a technique through which a reader's eye movements are measured so that the researcher knows exactly where the reader is looking and in which sequence the eyes are moving from one location to another. Eye movement research allows the researcher to gain insight into the readers' behaviour and hence a greater insight into their probable cognitive processing (Bax &Weir, 2012). Study One answered the Research Question: Do the eye movements of Arabic L1 EFL learners differ from those of skilled English L1 readers. If so, how do the eye movements differ and to what extent? Fixations, forward saccades and regressions were investigated using data from 11 eye movement metrics. Specifically, these metrics were 1) number of fixations 2) mean fixation duration 3) forward saccade length 4) number of forward saccades per individual 5) length of regressions and 6) proportion of saccades that area regressions 7) sum of all visits on consonants 8) mean of all visits on consonants 9) sum of all visits on vowels 10) mean of all visits on vowels and 11) mean proportion of visits on vowels. The study aimed to investigate the similarities and

differences of the two groups in order to better understand the difficulties experienced by the Arabic L1 EFL students when reading in English.

Study Two was carried out to ascertain if focused pedagogical interventions could help to develop word level reading skills which would be manifested in the eye movements of the Arabic L1 EFL participants. Specifically, these interventions included textual enhancement, tracking activities, the explicit learning of words included in New General Service Word lists, rapid word recognition exercises and oral text reading fluency activities (see Chapter 5.5 for a detailed description of these interventions).

Study Two also investigated whether these interventions had any significant effect on the overall reading comprehension of the participants as assessed by a pre- and post-reading instrument. Study Two asked the following Research Questions: 1) Do focused pedagogical interventions change the eye movements of Arabic L1 EFL students while reading at the word and sentence level? If so, how, and to what extent? 2) Do focused pedagogical interventions influence word and sentence level reading processes and thus overall reading comprehension? If so, how, and to what extent?

1.4 Structure of the thesis

This thesis consists of six chapters including the Introduction. In Chapter 2, a detailed discussion of theoretical concepts related to the two studies is provided. This chapter reviews key concepts such as the cognitive processes in both first and second language reading. It discusses the nature of the Arabic language and the difficulties experienced by Arabic L1 EFL students. Following this, basic concepts and terminology of eye movement research are discussed. Chapter 3 of this thesis presents a thorough discussion of the pedagogical interventions of textual enhancement, phonological awareness training, training in word

recognition and automaticity, spelling instruction and oral text reading fluency. The theoretical concepts of noticing, attention and awareness are also discussed. Chapter 4 contains the rationale for and a description of the research design of Study One, a quasi-experimental study which compared the eye movements of skilled English L1 readers and Arabic L1 EFL learners. Participants, materials, procedure, data analysis and results are described in detail. Chapter 5 presents the rationale for and a description of Study Two, a study which was carried out to ascertain if focused pedagogical interventions could help to develop word level reading skills which would be manifested in the Arabic EFL students' eye movements. Study Two also investigated whether the pedagogical interventions had a significant effect on the overall reading comprehension of the participants as measured by a pre-and post-reading test. Participants, materials, procedure, data analysis and results are described. Chapter 6 provides a summary of the main findings of Study One and Study Two and outlines the contribution the two studies make to the literature. It also discusses the pedagogical implications of the results and the limitations of the studies. Chapter 6 concludes with suggestions for further research.

Chapter 2: An overview of reading processes and the use of eye tracking in researching reading

Introduction

In this chapter, I will discuss the cognitive processes in first and second language reading and will then describe the cross-linguistic approach to reading. This will be followed by a description of the nature of the Arabic language, the reading patterns of Arabic L1 speakers and the difficulties experienced by these students when reading in English. The next section of this chapter deals with the basic concepts and terminology of eye tracking, a historical overview of eye movement research and a synopsis of major findings from eye movement studies.

2.1.1 Cognitive processes in first language reading

Many of the current views on L2 reading have been influenced and shaped by research on L1 reading. Although there are significant differences between the two, it seems a legitimate starting point to consider what first language research has discovered about the nature of the cognitive processes in reading and the development of fluent reading ability in the L1. Reading is a complex activity and there is still no broad theory of reading that can explain its full width and complexity because it has too many components for a single theory (Perfetti & Stafura, 2014). Grabe (1991) has pointed out that "simple definitions typically misrepresent complex cognitive processes such as reading" (p. 378). Despite this caveat, the much-cited Simple View of Reading (SVR) proposed by Gough and Tunmer (1986) and Hoover and Gough (1990) suggests that reading comprehension is the product of two basic components: *word reading* and *language* (listening) *comprehension*. However, this simple definition does not imply that reading, or learning to read, is a simple process. On the contrary, it is "a simple way of conceptualising the complexity of reading" (Oakhill, Cain & Elbro, 2015, p. 2).

Word reading (or word decoding) is the ability to read single words out of context, either by letter-sound associations or by recognition of a unique letter sequence. That is to say, the reader has two different strategies available for identifying a printed word: non-lexical and lexical. The non-lexical, or phonemic strategy, involves translating the word into an auditory representation and then using this representation to retrieve the meaning of the word. The lexical, or visual strategy, involves using visual information to retrieve the meaning (see Coltheart et al., 1993, 2001, 2005; McClelland & Rumelhart, 1981, 1986 for a detailed discussion of this 'dual-route theory of reading). Language comprehension is the ability to understand words, sentences and texts in order to build a mental representation of the content of a text. The two components, although separate skills, work together and are both necessary for text comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990). However, they do not develop in tandem. By the time a child enters school, language comprehension is relatively well-developed (Cain & Oakhill, 2014). Word decoding, on the other hand, is a new skill which will need to be learned. The child will need to learn how to decode words fluently and automatically, a skill which, although not sufficient in itself, is necessary for successful reading comprehension.

In alphabetic languages, children need to be taught the relationships between the sounds and the letters, or letter combinations, which represent these sounds i.e. the grapheme–phoneme relations. This association of letters with phonemes is referred to as the 'alphabetic principle'. This includes phonemic awareness, which refers to the knowledge that sentences are made up of words, which are made up of groups of sounds, which are comprised of phonemes. Without adequate phonemic awareness, learners are unable to blend sounds together to form words or to segment words into their individual sounds. Phonemic awareness is now accepted as not only a strong predictor of future reading ability, but also a primary cause of word-level reading difficulties

(Howard, 2002). An understanding of these graphophonemic relations must be obtained through either explicit instruction or implicit learning and practice. This is not an easy task, as young children have an imperfect idea of what phonemes are because these are abstractions (Rayner, Foorman, Pesetsky & Seidenberg, 2001). However, in time, children learn about spelling patterns that recur in different words and these larger units are used to form connections to remember words (Bhattacharya & Ehri, 2004). When readers acquire sufficient knowledge of the alphabetic principle, they are able to learn sight words quickly and to remember them long term.

Another component of word recognition is orthographic processing which is based on the formation of visual long-term memory representations of letters, letter patterns, and sequences of letters that serve to spatially 'map' the temporal sequence of phonemes within words (Ehri, 1992;1998; 2005). Children acquire orthographic processing skills with repeated exposure to printed words, which enables them to develop stable visual representations of letter sequences, word parts and whole words in long-term memory (Barker, Torgesen &Wagner 1992). Developing accurate visual representations of letters and words in long-term memory is essential for rapid word recognition and proficient spelling (Ehri 1991, 1992). By repeatedly associating a word's correct pronunciation with its visual representation, readers develop a memory for spelling patterns that are larger than individual letters. Orthographic development is often attributed to reading experience in which repeated exposure to print provides the foundation for this development (Rayner, 1998). However, not all readers benefit from implicit learning with practice. Several studies report that poor readers require more exposures to learn novel words than skilled readers. On average, poor readers needed to be shown a novel word 9.2 times to recognize it, in comparison with the 6.8 exposures that good readers needed (O'Brien et al., 2011). Cumulative research in this area indicates that exposure to words affords different experiences for different levels of

readers. The contribution of orthographic processing skills to reading depends upon the nature of the task and age of the reader. The relative contribution of orthographic processing to reading skill increases with age, and evidence suggests that individual differences in orthographic processing skills also predict response to reading intervention. In a study with second and third graders, Foorman, Francis, Fletcher, Schatschneider and Mehta (1998) found that efficiency of word-decoding was predicted by children's prior levels of orthographic knowledge.

Eventually, the child will attain a level at which word decoding becomes what Stanovitch (1995) terms a 'self-teaching mechanism'. In other words, with sufficient practice, the child learns to decode an increasingly large number of words without further teacher feedback (Ziegler, Perry & Zorzi, 2014). Once children start reading, most new vocabulary is learned through reading, not from being taught (Cunningham, 2005; Nagy & Scott, 2000). Thus, vocabulary supports reading comprehension and "reading with good comprehension supports vocabulary development, meaning that there is reciprocity between the development of these competencies" (Oakhill, Cain & Elbro 2015 p. 60). Some researchers regard word reading as the only skill, other than language comprehension, required to understand a written text (Gough & Tunmer, 1986; Hoover & Gough, 1990). The Simple View formula presented by Gough and Tunmer (op.cit) is: Decoding (D) x Language Comprehension (LC) = Reading Comprehension (RC). It has been suggested, however, that the relationship between decoding and reading comprehension decreases with increasing chronological age and language comprehension becomes a better predictor of reading comprehension (García & Cain, 2014).

Text comprehension is a complex task that necessitates many different cognitive skills and processes (Cain, Oakhill & Bryant, 2004). Taxonomies of comprehension abilities often categorize the component skills and processes as higher- or lower-level in the language processing chain. For

example, word recognition is considered a lower-level processing skill, whereas inferencing is considered a higher-level processing skill because it aids in the construction of a meaning-based representation of the text (Pressley, 2000). Readers do not remember the wording of a text; they remember the meaning and derive an overall representation, or mental model, of the text (Johnson-Laird, 1983). *Figure 2.1* below from Oakhill et al. (2015) shows an overview of some of the skills and processes involved in building an appropriate mental model of the text.

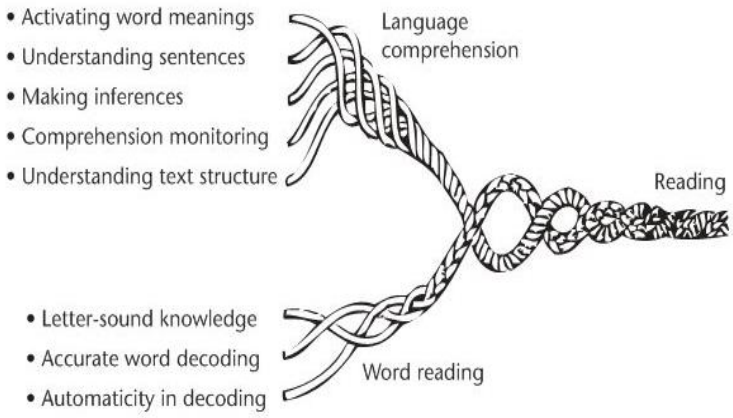


Figure 2.1 An overview of some of the component processes of reading

As can be seen from *Figure 2.1*, there are five strands involved in language comprehension and three in word reading, which when braided together, result in efficient and fluent reading comprehension. A summary of these component processes is as follows:

1) Activating word meanings. As mentioned previously, word decoding skills are necessary but not sufficient for successful reading comprehension. Additionally, the child will need to know the meanings of the words he or she encounters on a page. However, there are degrees to which one can know the meaning of a word and it is this 'depth' of word knowledge which is important for building mental models of text (Oakhill et al., 2015). In addition, fast and

accurate access to word meanings is crucial; therefore, speed of activation is an additional requirement for adequate comprehension of the text.

2) Understanding and linking sentences. Writing is generally more complex than spoken utterances and the child must learn how all the words are related and how the sentences are linked together in a coherent written text. If the child is to fully comprehend a written text, cohesive devices such as connectives and anaphoric reference will also need to be learnt and understood.

3) Inference making. Not all information in a written text is explicitly stated and the construction of a meaning-based representation involves generating inferences (Cain, Oakhill & Bryant, 2004). The developing readers will need to make inferences using background knowledge while reading, in order to understand what is happening on the page they are reading when the information is not explicit.

4) Comprehension monitoring. An efficient reader keeps track of his/her comprehension and, when inconsistencies appear, can try to remedy the problem. Rereading a word or sentence or using background knowledge to interpret the break in comprehension are strategies that a fluent reader uses.

5) Understanding text structure. Comprehension requires understanding text structure because it can help the reader identify the main idea and provide a framework for the mental model (Cain, Oakhill & Bryant et al., op.cit.). The structure varies according to the genre of the text and, as the beginning reader is exposed to various text types, he or she will need to learn how these different texts are organized. Best, Floyd and McNamara (2008) suggested that comprehension of a narrative text is more strongly influenced by decoding skills, while for expository text, word knowledge is a far better predictor of reading comprehension than decoding skills. Perfetti (1984) proposed that inadequate knowledge of text structure is a possible source of comprehension failure.

The ‘lower’ strands in the reading tapestry include letter-sound knowledge, accurate word decoding and automaticity in decoding. Ehri (1991) distinguished and described four different ways to read words. The first way is by decoding. The reader sounds out and ‘translates’ graphemes into phonemes, or works with larger groups of letters to blend syllabic units into words that he or she can recognise. The second way is by analogizing (Goswami 1986). This process involves using words which the readers already know in order to help them read a new word. Another way is by prediction (Goodman, 1970; Tunmer & Chapman, 1998), in the process of which readers use context and letter clues to guess unfamiliar words. These three ways of reading assist in decoding words which are unfamiliar. The fourth way of reading explains how readers decode words they have already seen before. When readers’ eyes land on a word known by sight, the word’s identity is activated in memory very rapidly (Ehri, 2005). When sight words are known well enough, readers can recognize their pronunciations and meanings automatically without any attention or effort at sounding out letters (LaBerge & Samuels, 1974). Despite popular misconception, the term *sight word* does not refer only to high-frequency words. Any word that is read sufficiently often becomes a sight word that can be recognized using memory representations (Ehri, 2005). Another misunderstanding which Ehri (op.cit.) mentioned is to consider sight word reading as a strategy. She explained that being strategic involves choice. Readers are strategic when they figure out unknown words by decoding, analogizing, or predicting. However, they are not being strategic when they read words by sight, which happens automatically and is not a matter of choice (Ehri op.cit. p. 170). The first three ways of reading a word require conscious attention, whereas sight word reading is an unconscious automatic retrieval process. When decoding a text, the reader’s attention shifts from the text to the word itself in order to identify it and this might disrupt comprehension. When word-level processing is inadequate, reading speed and

comprehension are negatively affected. Therefore, automaticity in decoding is essential for efficient and fluent reading.

All the component processes involved in reading comprehension require that the information (either from word or text knowledge) is both available and accessible (Cain, Oakhill & Bryant, 2004). Working memory is a temporary storage and processing system that is necessary for a range of cognitive tasks (Baddeley, 1986; Daneman & Carpenter, 1980). Working memory holds the most recently read content and also holds information retrieved from long-term memory to facilitate its integration with the currently active text. It is a resource that affects an individual's ability to carry out many of the processes associated with the construction of text representation. Slow or inaccurate word reading is proposed to affect comprehension by using up too much processing capacity with little remaining for text comprehension processes such as integration and inference (Hannon & Daneman, 2001; Perfetti, 1985). Working memory, which has three components: the phonological loop, visual-spatial sketchpad and the central executive (Baddeley & Hitch, 1974) is related to many aspects of reading comprehension. The phonological loop is dedicated to the temporal storage of verbal information and the central executive coordinates the storage and processing of incoming information. A meta-analysis of 77 studies in English L1 contexts conducted by Daneman and Merilke (1996) confirmed a strong relationship between working memory and reading comprehension abilities. Cain, Oakhill and Bryant (2003) established that working memory capacity explained unique variance in reading comprehension of readers between the ages of 8–11 years, after the contributions made by word reading skill and verbal ability had been considered. Working memory generally accounts for reading comprehension abilities above and beyond a wide range of specific skills, such as inference making, comprehension monitoring and understanding text structure (Cain & Oakhill 2006).

Recently, much discussion has centred on the role of working memory capacity as a source of individual differences in reading abilities of readers, otherwise matched for age and education (Baddeley, 2007; Cain, 2006).

2.1.2 Cognitive processes in second language reading

As was mentioned previously, there are several similarities between L1 and L2 cognitive processes in reading, and researchers from both domains agree that reading is not a unitary process. Grabe (2009) divided the cognitive processes involved in L2 reading into two levels: lower-level and higher-level cognitive processes. The lower-level processes involve such operations as letter identification, word recognition, syntactic parsing and proposition encoding. The higher-level processes include integrating information within a text, activating and utilizing background knowledge and making inferences (Khalifa & Weir, 2009; Yamashita, 2013). What distinguishes the lower from the higher-level processes is not their level of simplicity or complexity but rather the fact that lower-level processes can become strongly automatized while the higher-level processes require attentional resources (Field, 2004; Grabe, 2009; Khalifa & Weir, op.cit.).

Researchers from various disciplines agree that working memory has a limited capacity (Baddeley, 1986, 2007; Daneman & Carpenter, 1980; Ellis, 2001; Just & Carpenter, 1980; Oakhill, 1984). If a reader's attention is consumed with the lower-level processes of text decoding, there will be fewer resources available within the working memory for the use of general knowledge about the wider context. Lower-level processes inform the higher-level processes; therefore, if the former are slow and laborious, passage-level reading comprehension may not be successful.

Word recognition is one of the most important and well researched components of the lower-level processes. "Individual words are the critical building blocks in text-meaning

construction, and efficiency in converting graphic symbols into sound or meaning information is indispensable in comprehension" (Koda, 2004, p. 29). However, research notably in corpus linguistics has demonstrated that a large amount of communication makes use of fixed expressions memorized as formulaic chunks. Sinclair (2005) argued that the unit of language to be applied is "the phrase, the whole phrase and nothing but the phrase" (p. 110). Erman and Warren (2000) estimated that about half fluent native text is constructed according to the idiom principle and Ellis (2012 b) defined language learning as, "in essence, the learning of formulaic sequences" (p. 17). These formulaic sequences play a central role in both first and second language acquisition and fluent language users have a vast repertoire of memorized language sequences (Ellis, 1996; Ellis & Cadierno, 2009; Granger & Meunier, 2008). However, my research is primarily concerned with language at the word level and, as such, emphasis will be on word recognition.

Word recognition is considered to be an interactive process of accumulating phonological, orthographical, semantic, and possibly syntactic and morphological information (Harrison & Krol, 2007) and is widely accepted by researchers as one of the most important processes contributing to L2 reading comprehension (Grabe, 2009). To better understand the complex processes involved in L2 reading, numerous researchers have employed componential analysis (see for example Bernhardt, 2000; Hacquebord, 1989; Nassaji & Geva, 1999; Schoonen, Hulstjin & Bossers, 1998). In such analyses, "reading comprehension performance is explained by the performance on tasks assumed to measure constituent components of the reading comprehension process" (Van Gelderen, Schoonen, Stoel, deGlopper & Hulstjin, 2007, p. 477).

Khalifa and Weir (2009) proposed a cognitive processing model for reading comprehension. *Figure 2.2* below illustrates their model of reading. The *goal setter* will select the appropriate type of reading required for specific texts while the *monito*r is activated corresponding

to the goals of the reader. The options available to the reader as presented by Urquhart and Weir (1998) are global versus local and careful versus expeditious reading. Global refers to the comprehension of information beyond the sentence level. It involves incorporating information from different parts of the text. Local, in contrast, refers to comprehension at the sentence level and involves word recognition, lexical access and syntactic parsing. Careful reading denotes reading a text to obtain complete meaning at either the global or local level. Expeditious reading, on the other hand, indicates reading which is selective and quick in order to retrieve specific information. Khalifa and Weir (2009) make a distinction between three different kinds of expeditious reading: skimming, scanning and search reading. They define skimming as reading quickly to obtain the gist or main idea. Scanning occurs at the local level and involves the reader attempting to locate specific predetermined words, phrases or dates. Search reading also involves predetermined topics; however instead of exact word matches, the reader is looking for words in the same semantic field as the required information. Search reading can occur at both the local and global level. Khalifa and Weir (op.cit) argued that in addition to the level of processing required, reading proficiency is also a function of the complexity of the text and task the reading is carried out on.

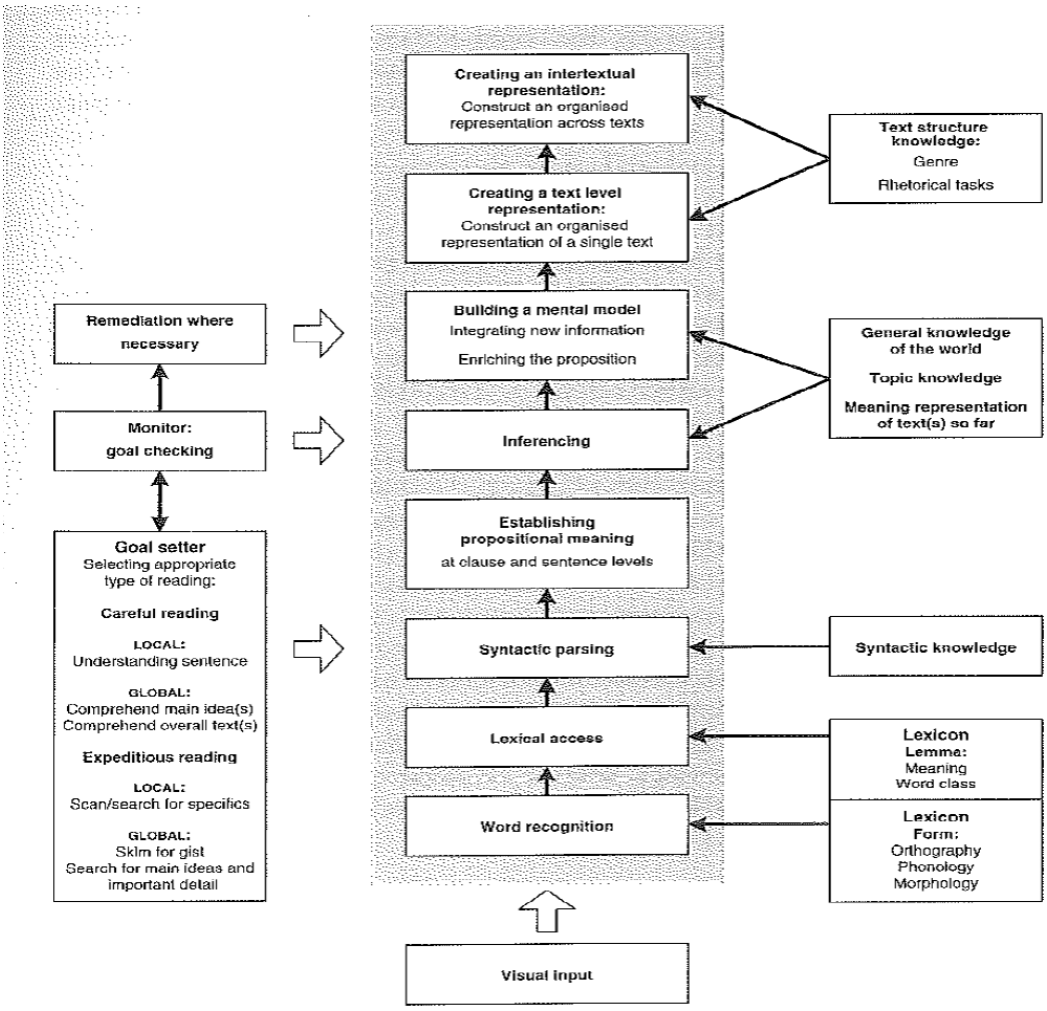


Figure 2.2 Khalifa and Weir's (2009) cognitive processing model for reading

Once the goal setter has determined the specific type of reading to be employed, vital decisions are made by the reader which affects the levels of processing to be activated. These processes are found in the central core of the model. *Word recognition* means matching the form of a word with a mental representation of the orthographic form. As mentioned in 1.2, Coltheart's (1978) dual route theory proposed that the reader processes written words by either a lexical or a sub-lexical route. *Lexical access* is the process of retrieving information about a word's form and meaning from the mental lexicon. Once lexical access is accomplished, the reader groups words

into phrases and clauses to understand the meaning. This is known as *syntactic parsing* and is related to word order, morphology and parts of speech. *Establishing propositional meaning at the clause or sentence level* is the understanding of the written words, devoid of the reader's external knowledge which would situate the words in a specific context. *Inferencing* is the process whereby the reader adds information which is not explicitly stated in the text to make meaning from the words. *Building a mental model* is the process whereby new information is incorporated into previous material. This process involves the ability to identify main ideas and connect them with prior ideas. It is conditional and likely to be revised as new information is incorporated. At this level, information may be categorised into what is important and what is less relevant. *Creating a text level representation* is the means by which the reader identifies the hierarchical structure of the text as a whole and decides which elements of information are central to text comprehension. *Intertextual representation* involves choosing and connecting information from more than one text.

Although the differences between first and second language reading are more obvious with beginning and weak readers, there are also important differences between first language readers and second language readers at an advanced level (Grabe, 2009). L2 readers do not have the luxury of waiting several years to build language knowledge before they begin to read, so they must develop linguistic resources at the same time that they develop reading comprehension (Grabe, op.cit.). Second language readers are working with the resources of two languages (Bialystok, 2001) and so it is likely that they will engage in different types of metalinguistic processing while reading in the L2. There also exist processing differences associated with practice in reading and the visual information in the orthography of the L2. L2 readers will be slower in word recognition and less accurate in reading in the L2 (Grabe, op.cit.). A final difference is the amount of exposure

to print that a learner experiences. Most L2 readers have limited exposure to L2 print most of which comes from L2 classroom contexts.

2.1.3 Cross-linguistic approach to reading

The central assumption underlying the cross-linguistic approach to reading is that "L1 experience embeds habits of mind, instilling specific processing mechanisms which frequently kick in during L2 reading" (Koda, 2004, p. 9). Hamada and Koda (2008) proposed that not only is the L1 activated during L2 reading, but that well-established L1 reading abilities are activated automatically by L2 print input, regardless of the learner's intent. L2 reading, unlike L1 reading, involves two languages and the dual-language involvement implies continual interactions between the two languages, as well as constant adjustments in accommodating the conflicting demands each language imposes (Koda, 2007). The 'reading universals hypothesis' (Goodman, 1973) proposed that the reading process will be similar for all languages. This view of reading in a foreign language was supported by other early researchers (see for example Coady, 1979; Rigg, 1977). Clarke (1979) stated if the reading process is fundamentally the same in all languages, we should logically expect good first language readers to be good second language readers. Several studies carried out in Canada with bilingual readers (Barik & Swain, 1975; Lambert, 1975; MacNamara, 1970) provided support for this hypothesis. However, these studies were later criticized by Cummins (1976) on the grounds that 1) there is no one single phenomenon called bilingualism, 2) the researchers did not control for participants' proficiency in the first and second language, 3) some studies used 'balanced' bilinguals and 4) the issue of additive or subtractive bilingualism was not addressed. Further weakening the claims of the Canadian studies is Alderson's (1978, 1979,) argument that the cloze test, which was used in the above cited studies to measure reading

comprehension is a measure of lower-level language ability, rather than of higher-order reading ability.

Cowan and Sarmad (1976) conducted a study with bilingual English-Farsi children which also contradicted the findings of the Canadian studies with bilingual readers. However, Cowan and Sarmad (op.cit.) proposed a theory of reading which explained the discrepancy. Their 'parallel processing theory of reading' posits that bilingual children have two sets of processing strategies and that transfer will be facilitated according to the similarity of the two languages involved i.e. the greater the difference between the languages, the less likely the learner is to read the first language in the same way as he or she reads the foreign language. Therefore, it is less likely that there will be a positive effect of transfer from the first language to the second language. In that same year, Alderson, Bastien and Madrazo (1976) offered more evidence that a learner's knowledge of a second language is more important to the comprehension of L2 texts than is reading ability in the first language. This study, conducted in a Mexican university, concluded that the best predictor of reading ability in L2 was not reading ability in the mother tongue, but rather proficiency in the second language. Clarke (1979) studied the relationship between L1 and L2 reading ability in the same individual. He suggested that there might exist a 'language competence ceiling' which impedes the proficient L1 reader from using effective reading behaviours in the L2. Cummins (1979) supported Clarke's 'short circuit' hypothesis and purported that there is a threshold level of linguistic competence which bilinguals need to achieve before the supposed benefits of bilingualism can appear. Since the attainment of these thresholds is determined by social, attitudinal, educational, and cognitive factors combined, these thresholds are "an intervening rather than a basic causal variable, in accounting for the cognitive growth of bilinguals" (Cummins, 1976, p. 23). In reviewing the evidence of over a decade of research

Alderson (1984), in his seminal article, asked the question "Is reading in a foreign language a language problem or a reading problem?" His answer was that it appears to be both, with strong evidence that it is more of a language problem for low levels of foreign language competence than a reading problem**.** Data from a longitudinal study of Dutch L1 and English L2 high school students conducted by Van Gelderen et al. (2004) supported the assumption that processing efficiency on both word and syntactic level are important for L2 text comprehension and also confirmed that linguistic knowledge (mainly L2 vocabulary knowledge) is a key predictor of successful reading in the second language as Alderson (op.cit.) maintained.

Grabe (2009) suggested that among the differences that influence how L2 readers process a text are the various orthographies that visually represent the phonological and morphological systems of each language. Writing systems differ on two dimensions: orthographic representation and depth. Orthographic representation refers to the linguistic unit each graphic symbol denotes. Orthographic Depth (Katz & Frost, 1992) refers to the degree of sound-symbol correspondences. An orthography that closely represents the phonology in a clear one-to–one relationship is called a shallow or transparent orthography. An orthography that does not closely represent the phonology is called deep or opaque. With regards to Arabic, which is the L1 of the EFL participants of my study, vowelized Arabic is considered to be a shallow orthography while unvowelized Arabic is described as a deep orthography (Abu-Rabia & Siegel, 2003). English is often considered as having one of the most opaque orthographies in the world. The distance between English and Arabic in terms of both depth and orthographic representation has been noted by numerous researchers (see for example Fender, 2003, 2008; Hayes-Harb, 2006; Randall & Meara (1988); Ryan & Meara, 1991).

2.1.4 Nature of the Arabic language

Arabic belongs to a group of languages called Semitic languages. However, unlike other Semitic languages, spoken Arabic has different dialects which vary from one geographical area to another. The difference between these vernaculars is evidenced at the phonological, morphological, syntactic and semantic levels (Taha, 2013). Each of the different variations also differs from Modern Standard Arabic (MSA), which is the literary language for all Arab speakers. Generally speaking, the first time Arabic speaking children are exposed to Modern Standard Arabic is when they begin to read and write. According to Saiegh-Haddad (2003a, 2003b), the linguistic distance between the spoken dialectical variations and MSA could be the main reason for a delay in the development of phonological awareness among many Arabic speaking children.

In general, most words in the Arabic language are morphologically derived from roots. The root presents the basic and the general semantic meanings of all the words which are derived from it (Taha, 2013). Most of these roots consist of three consonants, and these consonants can be combined with different vowel patterns to produce a whole family of words that share a common meaning. For example, the root **k-t-b** (which has the basic meaning of marking, inscribing or writing) combines with vowel patterns to produce katîb (writer), maktaba (library or bookstore), maktab (office), ketaab (book), iktitâb (subscription), kataba (he wrote) and so on. In Modern Standard Arabic (MSA) in addition to two diphthongs, there are three pairs of vowels, the vowels in each pair being distinguished by length. The long variants are the only ones always represented in writing. The shorter variants are not written, rather they are indicated only in children's books and in the Holy Qur'an by diacritical markers placed above the consonant that precedes the vowel sound. As only the consonants are written down, the reader is required to fill in the vowels which

are appropriate to the context. For example, it would not be possible to know the exact meaning of كتب as it could mean either *he writes* or *it was written.*

Reading is therefore "highly context dependent because there is no strategy to identify words that are visually identical and carry different meanings" (Abu-Rabia, 1995, p. 6). "Arabic is perhaps the only language in the world in which readers must first understand the sentence in order to recognize the word" (Abu-Rabia, 1997, p. 76). In contrast, English words with similar consonant structures are not always semantically related. Therefore, the reader of English cannot predict the meaning of a word based on its consonant structure alone. When reading in English, Arabic L1 speakers who rely on the decoding system of their L1, assume that words like *red, read, raid* and *ride* are related (Hayes-Harb, 2006).

It is not uncommon for syllables and even entire words to be spelled without vowels. As illustrated in *Figure 2.3,* the first three words have a syllable without a vowel, while the last three words are spelled completely without vowels.

Word	**Written Arabic**	**English Transliteration**
ketaab (book)	كتاب	k t a b
katîb (writer)	كاتِب	k a t b
iktitâb (subscription)	اكتتاب	i k t t a b
maktaba (library)	مكتبة	m k t b h
maktab (office)	مكتب	m k t b
kataba (he wrote)	كتب	k t b

Figure 2.3 k-t-b root words

Skilled Arabic readers have learned to use an orthography that does not include diacritic markers to signal short vowels. Therefore, Arabic readers tend to focus their attention on the consonants (Hayes-Harb, 2006; Ryan, 1997; Ryan & Meara,1991; Thompson-Panos & Thomas-Ruzic, 1983). This strategy is quite different from that used by skilled English L1 readers who process vowels (see Ashby, Treiman, Kessler and Rayner (2006) for a discussion of vowel processing during silent reading).

2.1.5 Reading patterns of Arabic L1 speakers

Katz and Frost (1992) observed that the reading process varies for different orthographies. The array scanning studies of Randall and Meara (1988) demonstrate that reading patterns may, indeed, not be universal. They investigated the ways in which native speakers of Roman alphabets (RAs) and native speakers of Arabic process strings of letters and digits in English and Arabic. Array scanning involves measuring the speed by which different letters or shapes presented in arrays can be detected by participants. In their study, participants were shown a target letter or digit, on a computer screen which was then removed. Participants were then presented with a five-item array and their task was to say as quickly as possible whether the target letter or digit appeared in the set of five or not. The Roman letters consisted of a set of 25 upper case computer generated letters on a dot matrix which was 8 rows deep and 8 columns wide. Arabic letters were generated on the same matrix (see *Figure 2.4* below for an example).

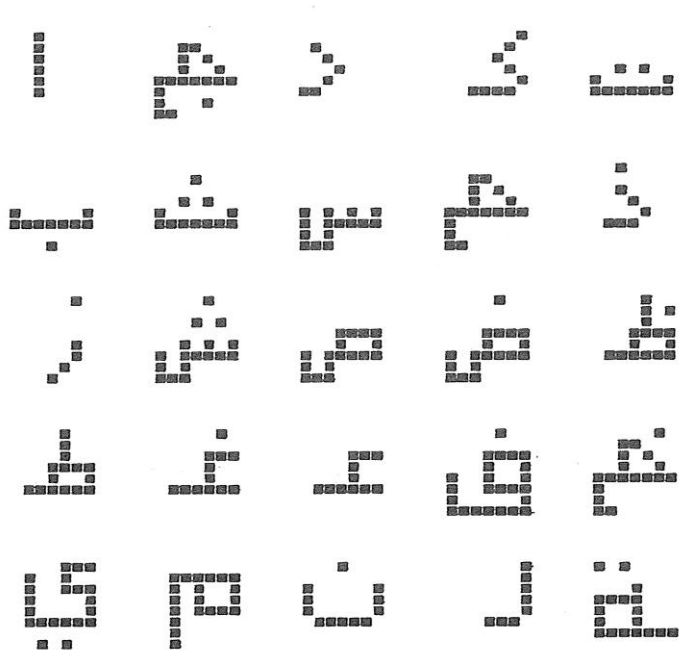


Figure 2.4 How Arabs read Roman Letters (Randall & Meara, 1988)

Reaction times for correct responses were recorded. They found that the RAs processed arrays of letters and digits differently from the way in which they processed strings of shapes or letter-like forms which were unfamiliar. For non-alphabetic shapes, targets were recognized fastest if they were positioned in the middle of the array and more slowly if they were at the ends of the array. This pattern of response showed a characteristic U-shape. However, results for letter and digit searching were different. Here, the RAs produced the fastest response to targets on the left of the string and gave slightly faster responses to targets in the middle and in the rightmost end of the string. The characteristic search function that emerged was described as a tilted M-shape. When the Arabic L1 participants scanned the arrays of Arabic letters, they did so in a right to left direction, and produced a U-shaped curve which is the typical curve produced by RAs when they search arrays of non-letter shapes. However, when the Arabic L1 readers scanned arrays of English letters, they "reacted to the Roman letters in the same way as they reacted to Arabic letters"

(Randall & Meara, op.cit. p. 134) i.e. in a U-shaped curve. They concluded that reading direction was an important factor in the way Arabic speakers search arrays. One interpretation of this might be the attentional factors governed by reading habits. Thus, readers of right to left languages tend to position their attention more to the left of the fixation since this is where they extract the next source of information (Farid & Grainger, 1996). The Arabic L1 reaction times became faster over the year-long study. However, there was no sign in the data that the students adopted the characteristic M- shaped curve produced by the RAs (for more details on eye-movement research, see the following section).

When participants were asked to identify specific letters in arrays of letters, Randall (2007) also found radically different visual search strategies exhibited by his English L1 and Arabic L1 participants. He proposed that one of the possible explanations for the different search strategies employed by his participants is the effect of saliency. English words are marked by spaces and consequently an L1 English reader isolates words by looking at the gaps. In order for the eye to jump the proper distance the reader must know that a word is simply a letter or a group of letters surrounded by a space. However, the Arabic script has spaces both within and between words. Although the gaps within words are not as large as the gaps between words, Arabic words are not as clearly delineated as they are in English (See *Figure 2.5* below). The eye movements of Arabic L1 learners possibly reflect this aspect of the Arabic language. Randall (op.cit) suggested that Arab readers may be attending to quite different salient features to extract words from the text than are English readers.

الـ جمـ لة هذه فـ ي الـ كـ لمات فـ ي ثـ غرات هناك أن سـ تلاحظ

Figure 2.5 Spaces within and between words

Translation: You will notice there are spaces within the words.

In sum, both Randall and Meara's (1988) and Randall's (2007) studies provide support for cross-linguistic interactions to word decoding and for the hypothesis that Arabic L1 speakers approach English words differently from native speakers of English. However, as Randall (2007) himself pointed out, one of the problems with such psychological experiments is that they have little face validity in that the task "is a long way from anything we would recognise as reading and the interpretation of such findings in terms of reading as we know it, is highly speculative" (p. 19).

Viewed collectively, the results from the aforementioned studies confirm that reading and word recognition processes may vary in different languages and this is amplified with users of different scriptal systems (Randall, 2007). It would appear that Arabic L1 learners of EFL use an L1 processing strategy which is potentially less effective in English word recognition. Consequently, these learners need explicit training in word recognition skills to access the meaning of the words, "instead of spending cognitive space and energy trying to decode the word" (Randall, op. cit., p. 12). The search patterns exhibited by the Arabic L1 EFL participants in Randall's (2007) study did not change with increasing fluency and exposure to English; confirming research by Geva and Siegel (2000) who observed the independence of oral English proficiency from word-level reading skills. Cross-linguistic comparisons of L2 word recognition demonstrate that L1 orthographic experience has "long-lasting clearly detectable impacts on L2 lexical processing, further implying that L1 processing experience is a major source of performance variation among L2 learners" (Koda, 2004, p. 46). Clearly, the importance of word recognition strategies cannot be underestimated in the teaching of Arabic L1 students.

2.1.6 Difficulties experienced by Arabic L1 speakers reading in English

Numerous studies have been published on the difficulties exhibited by native speakers of Arabic when reading in English (for example, Abu-Rabia, 1997b; Fender, 2003, 2008; Hayes-Harb, 2006; Randall & Meara, 1988; Randall, 2007; Randall & Groom, 2009; Ryan, 1997; Ryan & Meara, 1991; Saigh & Schmitt, 2012; Thompson-Panos & Thomas-Ruzic, 1983). In an early study by Ryan and Meara (1991), proficiency matched participants were shown a word on a computer screen for one second and then it was blanked out. The word reappeared two seconds later, either spelled correctly or in an altered form. The altered form consisted of spelling with one vowel removed (i.e. dparatment, expriment, managment, sufficint). The researchers found that performance by native Arabic speakers was less accurate and slower than performance by EFL learners with non-Arabic native language backgrounds or native English speakers. In particular, the native speakers of Arabic had a higher error rate in judging deleted vowel stimuli than the two other participant groups. The authors indicated that because of the lexical structure and orthography of their first language, Arabic L1 learners of English rely heavily on consonants when attempting to recognize an English word and they concentrate on the position of the consonants at the beginning, middle and end of words rather than on the position of vowels. In other words, their search functions are radically different from those readers whose script uses the Roman alphabet. This transfer from processing Arabic orthography impeded processing of English script in a phenomenon they called "vowel blindness" (Ryan & Meara, op.cit). Ryan (1997) stated that Arabic L1 speakers ignored the presence of vowels when storing vocabulary and made "an almost indiscriminate choice as to which vowel to use when one was needed" (p. 189). She further added that if the Arabic L1 speakers are unsure about the exact sounds of the vowels they hear in English, they may place excessive reliance on recognizing words by their consonants.

Hayes-Harb (2006) replicated Ryan and Meara's (1991) study but also included a third stimulus condition where consonants were deleted to serve as a control. By including this condition, it was possible to "determine whether native Arabic speakers are less sensitive to deleted vowels than the other two participant groups and whether they are more or less sensitive to deleted vowels relative to deleted consonants" (Hayes-Harb, 2006, p. 326). She hypothesized that the Arabic L1 learners of English would exhibit a pattern of attention to vowels and consonants which would differ from English L1 speakers, as observed by response time, accuracy rate and letter detection. She found that the Arabic group's responses were significantly slower than those of the non-Arabic groups. However, the Arabic group had approximately the same reaction times as the other two groups when responding to deleted vowels versus deleted consonants. The response accuracy data showed that the Arabic group's performance was significantly less accurate than that of the English group. In the letter detection test, the Arabic group exhibited the least accurate performance overall. She proposed that native speakers of Arabic are 'less aware' of vowel letters than the two control groups from different language groups. She did not provide an operational definition of 'sensitive' or 'aware'. However, she noted that in post-experiment conversations, participants revealed that they were consciously aware of the prominence of consonant information over vowel information in their English word identification processes.

Fender (2003, 2008) concluded that Arabic ESL students experience 'more difficulty' than other ESL populations he studied in processing English word forms. Results from his (2003) study found that native Arabic speakers were significantly slower than a group of proficiency-matched Japanese speakers in a lexical decision task and he argued that Arabic speakers have slower and less effective context-free word recognition skills. He maintained that difficulties acquiring

English spelling knowledge not only affect word recognition skills but also constrain reading skills (Fender, 2008).

Randall and Groom (2009) described the instances of consonant and vowel errors found in the British University in Dubai Learner Corpus (BUiD), a corpus of 287,227 word tokens and 20,275 word types:

> "The striking thing about the type of errors is that they almost always preserve the consonant structure of the target word. The vowels are often incorrect, but more importantly, they are often omitted, or turn up in the wrong place relative to the surrounding consonants" (p. 532).

Saigh and Schmitt (2012) investigated the word-level spelling skills of Arabic L1 learners of English. 40 target words with short vowels and 40 words with long vowels were chosen. Participants were shown 80 sentences in which the target words were embedded. They were asked to decide whether the words were spelled correctly or not. In the instances where participants ticked 'incorrect', they were asked to make the corrections. The researchers reported that Arabic L1 learners of English had more problems recognizing spelling errors in words with short vowels than in words with long vowels. The results also showed that participants were able to notice spelling errors in which the English vowel was missing (e.g. conclsion) somewhat better than when the vowel was represented by the wrong letter (e.g. imprave). They interpreted these findings as suggesting that "Arabic speakers process English short and long vowels in a way similar to their L1, which affects their ability to spell and recognize English words" (Saigh & Schmitt op.cit, p. 24). The authors suggested that their study moved our understanding beyond the general 'vowel blindness' phenomenon proposed by Randall and Meara (1988) and Ryan and Meara (1991) and

demonstrated that the type of vowel also matters i.e. short vowels are significantly more difficult for Arab speakers to recognize and recall in spelling than long vowels. "It appears that long vowels are more salient than short vowels, and the difference in the Arabic L1 system in which short vowels are either not written or indicated by diacritic marks seems a likely explanation" (Saigh & Schmitt, 2012, p. 31).

There is substantial evidence that L1 word recognition strategies develop differently across different orthographies which will, in turn, influence the development of English as a Second Language (ESL)[1] spelling knowledge. Figueredo (2006) reviewed 27 studies that investigated the influence of ESL learners' first language on the development on English spelling skills. The studies provided evidence that ESL learners rely on their knowledge of L1 phonological and sound-to-spelling correspondences, and this influences ESL spelling development either negatively or positively. The direction of influence depends on the degree of similarity between the two languages' phonological/orthographic systems. Where similarities exist, positive transfer may provide the ESL learners with a knowledge and skill advantage. Where differences exist, negative transfer may temporarily occur until English conventions are learned and consistently applied (Figueredo, op.cit). The recurrent hypothesis is that Arabic L1 learners of English approach English texts with strategies that reflect the nature of the Arabic orthographic system and that these strategies are potentially not conducive to efficient processing of English.

As previously mentioned, a particular difficulty pertains to the processing of vowels. Results from English L1 eye movement experiments have indicated that consonant information contributes more heavily than vowel information to the early phases of word recognition during

[1] The term ESL will be used throughout this thesis to denote English as a Second Language learners i.e. those students who are studying English in an English-speaking environment. The term EFL will be used to denote English as a Foreign Language learners and refers to those students who are studying English in an environment where English is not the language of that country.

silent reading (Ashby, Treiman, Kessler & Rayner, 2006). However, data suggest that skilled readers of English do include the vowel information in the early phonological representations (Ashby, 2006). If we acknowledge that letter recognition is an important subcomponent of word recognition, classroom instruction that focuses on automatic word-recognition ability should be a critical element of any reading programme designed for Arabic L1 students. If automaticity of word recognition is a major attribute of L2 reading behaviour, then ways of encouraging its development should be found. It is therefore important to provide opportunities for learners to *notice* those aspects of the English language they are unlikely to pay sufficient attention to without guidance (Schmidt, 2010).

.

2.2 The nature of eye tracking

2.2.1 Basic concepts and terminology

Eye movement research allows the researcher to gain insight into the readers' behaviour and hence a greater insight into their probable cognitive processing (Bax & Weir, 2012). Eye tracking refers to techniques which are used to record and measure eye movements. The methods by which this is accomplished, with varying degrees of sophistication, have been used as a tool in psychological reading research for over 100 years. In this section, I will discuss several aspects of eye movement studies. First, I will provide explanations of some basic terminology and concepts related to eye movement technology and research. Then, I will illustrate how eye movement research has been used over the years as a window into reading comprehension difficulties. Then, I will introduce some of the major findings of contemporary eye movement research.

Eye tracking is a technique through which an individual's eye movements are measured so that the researcher knows precisely where the participant is looking at any given time, and in which sequence the eyes are moving from one location to another. An eye tracking machine follows the

user's eye movements by reflecting infrared light onto the eye and then, using a geometrical model, determines the exact gaze point of the user. The use of eye tracking data in psychological research rests on the premise of an "eye-mind link" (Just & Carpenter, 1980). Under this assumption, overt attention and covert attention are closely linked. Cognitive processing is considered a major determinant of when and where the eyes move during complex task performances such as reading (Just & Carpenter, op.cit).

The human eye is not ideally designed for reading. The major disadvantage of the structure of the eye with regard to reading is that only a small part of what is printed on the page is in focus on the retina (Samuels, Rasinski & Hiebert, 2011). The retina of the eye contains two kinds of cells: cone cells and rod cells which have different reading functions. Cone cells provide the visual acuity that enables readers to see letters and words clearly, but they are not evenly distributed across the retina. Instead, they are concentrated in an area known as the fovea. There are about 10 million cone cells in the foveal area where vision is most acute. The fovea is concerned with processing details; words presented to locations removed from the fovea produce a marked drop in acuity and are more difficult to identify (Rayner & Sereno, 1994). The foveal region is the area that we think of as being "in focus" and includes 2 degrees of visual angle around the point of fixation, where 1 degree is equal to three or four letters, thus, approximately 6-8 letters are in focus at a time. The other cells found in the retina are called rod cells. These are located outside the fovea in the parafoveal region and serve a double purpose. The parafoveal region provides information that is useful for word recognition and for planning the distance to be moved for the saccadic jumps as the eye moves from point to point (Rayner & Sereno op.cit). The parafoveal region extends to about 15 to 20 letters. The third region to which the eye has access is called the peripheral region. It encompasses everything else in the visual field beyond the parafoveal region. Although the most

effective processing is reserved for that done in the fovea, some processing can also be accomplished for information in the parafoveal and peripheral vision (see Schotter, Angele & Rayner, 2006 for a comprehensive review).

Because of the 'shortcomings' of the structure of the eye, rapid eye movements are required to bring different parts of the text onto the small part of the retina which can see letters and words clearly. These very small, high velocity jumps are called *saccades*. The primary function of saccades is to bring a new section of text into foveal vision because reading based only on parafoveal or peripheral information is difficult to impossible (Rayner, 1988; Rayner & Bertera, 1979). Saccade latency (the time needed to encode the location of a target in the visual field and initiate an eye movement) is approximately 175-200 milliseconds (Rayner, Slowiaczek, Clifton & Bertera, 1983). Saccade duration (the amount of time that is takes to actually move the eyes) is a function of the distance moved. A 2-degree saccade, typical of reading, takes about 30 milliseconds (ms.) (Rayner 2009). The average length of a saccade is approximately 7-9 letter spaces for skilled readers of English who normally take about three to four saccadic movements per second, each lasting between 20 and 40 ms. Since vision is supressed during a saccade, new information is not encoded. It had generally been assumed that there is a near perfect binocular coordination during reading and that the eyes typically land on the same letter in a word. However, recent research has shown that that up to 40-50% of the time the eyes are on different letters (see Kirkby, Webster, Blythe & Liversedge, 2008 for a review).

Saccades of particular importance are called *regressions*. These are saccades that move backwards in the text. Regressions due to visuomotor and word identification processes are usually quite short and involve mostly inter-word regressions or regressions that land on the word immediately to the left of the launch site (in left-to-right languages). Regressions owing to higher-

level processing problems are usually longer and the length of the regressive saccades can cover multiple words (Carreiras & Clifton, 2004; Frazier & Rayner, 1982; Meseguer, Carreiras & Clifton, 2002). Saccadic regressions occur about approximately 10-15 % of the time, in skilled readers (Dussias, 2010; Rayner 2009; Rayner & Polletsek, 1989).

Saccades are separated by moments (or rather seconds) during which our eyes remain still. These eye *fixations* are times when the eyes are relatively stationary and reflect when information is being encoded, allowing readers to extract information from the text. It is during these pauses that the information contained in the eye fixation is taken up and fed to the brain for analysis and meaning. Eye fixations during L1 reading in English last approximately 200-250 ms. (Rayner, 2009).

An important question in reading research concerns the amount of information that readers acquire at each fixation. Classic experiments have attempted to control how much participants can see on each fixation. In the moving window paradigm (McConkie & Rayner, 1975) only the part of the visual field around the gaze location is displayed normally; the surrounding part of the visual field is altered (removed for visual scenes or replaced by chains of X in reading). The moving mask paradigm (Rayner & Bertera, 1979) is a reverse technique in comparison with the moving window paradigm. It dynamically obscures central vision (or replaces letters with X in reading), permitting only extrafoveal information use. In the boundary technique (Rayner, 1975; Balota, Pollatsek & Rayner, 1985; Miellet & Sparrow, 2004), a word is displayed to the right of where the eyes are fixated, but when the eyes go into motion, that word changes into another word. Researchers vary the relationship between the first word and the second word so they can be semantically, phonologically or orthographically related. Then the researchers try to infer what kind of information participants would be processing from that word before it comes into sharp

focus. Collectively, these experiments have confirmed that the *perceptual span*, or the region from which readers are able to acquire useful information, is quite small (see section 2.7.3 for details).

2.2.2 Historical overview of eye movement research

When discussing historical overviews of eye movement research, Paulson and Goodman (1999) state:

> "It is unfortunate when workers in the field come to know a body of research only from reports of it in the current literature. Relying on third-party interpretations alone can lead to widespread misunderstandings or misrepresentations of the original researcher's data and findings" (p. 11).

Therefore, what follows is a review of several of the original research articles on eye tracking and reading remediation. Unfortunately, several of the original research articles are not available and therefore, using several secondary sources was sometimes unavoidable.

Louis-Emile Javal is widely credited as being the first writer to use the term 'saccade' (from the French 'saccader' to twitch or jerk) to refer to rapid eye movements. However, there is some disagreement (Wade & Tatler, 2009; Wade, 2010) as to whether it was actually Javal or his colleague Lamare who observed in 1878 that a reader's eyes do not move smoothly across print but make a series of saccades or jumps. The observations, whether by Javal or Lamare, were characterised by a reliance on naked-eye observation of eye movement in the absence of technology. Nevertheless, with the acknowledgement that the eye stops at certain places along a line of print, came the basis for exploring the role of eye movements in reading (Huey 1908). In 1891, Landolt, a colleague of Javal at the Laboratoire d'ophtalmologie in Paris, reported that reading of a foreign language required more pauses than did reading in one's first language, as did

the reading of detached words, numbers and lists of proper nouns (as reported in Huey 1908 p. 19). This provided the first piece of evidence that the eyes do not proceed on a regular predetermined path, but the trajectory varies depending on the type of reading being carried out (Paulson & Goodman, 1999). In 1891, Ahrens fastened a small ivory cup to the cornea of the eye and sought by means of a rod attached to a cup to have the eye movements recorded on a smoked drum. A "plaster of Paris cup was later substituted for the ivory cup by Professor Delabarre of Harvard" (as reported in O'Brien 1922, p. 6). This technique was perfected by Huey who attached a plaster of Paris cup with a hole in the centre to the cornea of the participant's eye. The cup was attached to an aluminium pointer which responded to the slightest movement of the eye. As the participant read, the pointer traced the movement of the eye on a piece of paper. In addition to demonstrating that the eye regresses a small percentage of the time, his studies showed that 1) the first fixation in a line is frequently not at the first word but at the second or third; likewise, the final fixation is usually not on the last word 2) children make more frequent and generally longer pauses than adults 3) the reading range of the eye is limited, with the average reading span being ten letter spaces and 4) that more is read to the right of the fixation than to the left (Huey 1908).

Gray (1917) took photographic records of eye movements during reading and reported that the number of pauses per line was conditioned by the width of the perceptual span. Consequently, he hypothesized that training would enlarge the perceptual span which would result in a reduction of the number of fixations per line, with a corresponding increase in the speed of reading. He trained a fifth-grade pupil in rapid silent reading and found that in every test, there was a widening of the perceptual span and a "decided improvement in speed" (as reported in O'Brien 1922, p. 49). Schmidt (1917) reported more fixations per line for oral reading than for silent reading. He found more regressions in the reading of the elementary group of students than in the higher groups. He

also noted that regressive movements were a feature of slow readers and younger readers. This was corroborated by Pressey (1926) who also reported that good readers had fewer fixations and regressive movements per line than poor readers.

Generally speaking, the 1920s witnessed an increasing interest in eye movement studies as they applied to the teaching of literacy, especially to the teaching of reading and remediation of reading difficulties. O'Brien (1922) constructed classroom training activities in effective rapid silent reading, based upon the findings of his experiments with eye movements. He believed that rapid efficient reading is characterized by rather uniform, rhythmical movements of the eyes in contrast with the irregular movements and frequent regressions of the stumbling halting reader. However, he also proposed that eye movements were also affected by other factors, such as the size of the visual span, the simplicity or difficulty of the subject matter, the purpose for which the text is read and the ability to grasp meaning quickly. Results from his eye movement studies indicated that the reader differentiates between different types of reading and "evidently approached different reading problems with a different mental set" (O'Brien op. cit., p. 60). Buswell (1922, 1937) and Judd and Buswell (1922) photographed readers' eye movements and provided additional evidence that readers read differently in different circumstances.

O'Brien (1922) believed that reading which was chiefly dependent upon foveal vision would be slow and halting and therefore he devised exercises that influenced "the perceptual process that occurs in the fixation-pause" (op. cit., p. 127). In another of his studies, photographic records were taken of the eye movements of pupils with reading difficulties while they were engaged in silent reading. Records were taken before and after training. Initial records showed the number of fixations was excessively large and regressive movements were numerous. However,

two months of experimental training produced a marked increase in speed and a widening of the visual span.

There appears to have been a hiatus in eye movement studies in the 1940s and 1950s (Rayner & Pollatsek, 1989) and not until the 1960s do we see an emerging interest in technology as applied to the teaching of reading. In a study by Taylor (1971), students from grade one to grade twelve read 100-word passages silently. While the students read, eye cameras recorded the number of fixations for the passage. Taylor observed that the eyes skipped certain words and that the words which tend to be skipped were determined, in part, by word length. Short words, high frequency words and words that can be predicted from context were skipped. In addition, he found that less skilled readers made more backward regressive movements and the duration of each eye fixation was longer, which accounted in part for the slower reading speeds of the less skilled readers.

The 1970s was also when Keith Rayner started to change the way eye movement research was conducted. His approach "emphasized the importance of building theories bottom-up from a large body of solid, replicated findings" (Clifton et al. 2016, p. 4) and was concerned with understanding how readers performed on real-world tasks. Rayner believed that reading was primarily a question of obtaining information from the printed page. This belief diverged from the view that reading was a 'linguistic guessing game' (Goodman, 1970). Rayner was also interested in how eye movement studies could inform best practice in the teaching of reading (see Rayner, Foorman, Perfetti, Pesetsky and Siedenburg (2002) for an inclusive research-based overview of how a child's reading develops). He concurred with other research which demonstrated that children need to master the alphabetic principle in order to become proficient readers. He also recommended that direct instruction in phonics, rather than on whole word or whole language approaches be used by reading teachers.

Eye movement studies have been predominately conducted by cognitive psychologists and psycholinguists. However, the past 10 years has seen a growing involvement of second language researchers and testers to explore areas that had traditionally been investigated by using off-line research instruments such as judgment tasks, think-aloud protocols and interviews (Conklin & Pellicer-Sánchez, 2016). Bax (2012); Bax and Weir (2012); Brunfaut and McCray (2015); Conklin and Pellicer-Sánchez (2016); Godfroid, et al. (2015); Godfroid and Winke (2013); Indrarathne and Kormos (2016); Pellicer-Sánchez (2016); Smith, (2012) are some relevant examples of eye movement studies in applied linguistics. The analysis of eye-movements is a particularly useful tool for L2 acquisition researchers because it allows for the study of moment-by-moment processing decisions during natural, uninterrupted comprehension, and critically, without the need to rely on participants' strategic or metalinguistic responses (Rayner, 1998, 2009).

2.2. 3 Major findings from eye tracking research

Measurements taken of the duration and location of eye fixations have taught researchers a great deal about how people acquire information from the printed text, how they represent it and how they integrate it in the course of understanding a text (Rayner, 1998). There are several areas in which there appears to be a general consensus in the eye movement literature.

Determinants of fixation duration and length

Several variables have been shown to influence fixation times. The most robust findings are concerned with word frequency, word familiarity, word length, age of acquisition of the word, and ambiguity of the word. How long readers take to process a word is influenced by how frequent the word is in the language. The effect of word frequency has been observed consistently across a wide variety of contexts when other factors have been held constant. Readers spend more time

fixating on low-frequency words than on easy, high-frequency words (White, 2008). Rayner (1997) anecdotally noticed that readers looked longer at infrequent words and Just and Carpenter (1980) reported similar frequency effects. As frequency and word length are sometimes confounded, Rayner and Duffy (1986) and Inhoff and Rayner (1986) controlled for word length. Nevertheless, both studies showed a strong effect of word frequency on fixation times. However, it must be noted that an increase in gaze duration, as a function of the average length of preceding words, has been proven significant only for English (Pynte & Kennedy, 2006). Although two words may have the same frequency value, they may differ in familiarity. Whereas word frequency is usually determined from corpus counts, word familiarity is determined from rating norms in which participants rate how familiar they are with a given word. Effects of word familiarity on fixation time have been demonstrated in a number of studies (Juhasz & Rayner, 2003; Williams & Morris, 2004).

It is well known that word length influences adult skilled readers' eye movements (Just & Carpenter, 1980; Rayner, Sereno & Raney, 1996). For children, these effects are larger, in that children are relatively slower to process words than adults. This may be because longer words have more letters available in more visually degraded vision. However, another possibility might be that the longer fixation duration may be caused by children needing very detailed visual information about a word prior to initiating a saccade to leave that word (Reichle, Rayner & Pollatsek, 2003).

Age of acquisition of words also affects fixation duration. This is determined both by corpus counts and by subjective ratings. Juhasz and Rayner (2003, 2006) demonstrated that the age of acquisition effect tended to be stronger than even that of word frequency. In general, the number of meanings a word has also influences how long a reader will look at it (Sereno,

O'Donnell & Rayner, 2006). Similarly, words that are phonologically ambiguous (like 'wind') also yield differential fixation times (Rayner, Pollatsek & Binder, 1998). We also know that eye movements are influenced by textual and typographical variables. Print quality, length of line and amount of space between letters all influence processing (Dussias, 2010).

Research using the boundary paradigm has revealed that when readers have a valid preview of the word to the right of fixation, they spend less time fixating that word than when they do not have a valid preview. Accurate parafoveal previews have shown to increase the probability that words will be skipped during first pass reading (Angele, Tran & Rayner, 2011). The size of the preview benefit is typically 20 –30 ms. (Rayner, 1978, 1989). There is now considerable evidence (Rayner, 1998, 2009) that, while looking at word n (the fixated word), readers obtain useful information (preview benefit) from word n + 1 (the word to the right of fixation). However, readers typically do not get preview benefit from word + 2 (Rayner, Juhasz & Brown, 2007). If the prime word is orthographically similar to the target word, there is greater facilitation (Rayner, 2012).

The ability to move the eyes from left to right changes gradually allowing children to focus on the relevant part of the word and to accomplish more efficient information processing during fixations (Rayner & Pollatsek, 1989). When children start reading, their fixations tend to be quite long (over 350 ms. in first grade) and they tend to make as many as two to three fixations per word (depending on the length of the word). Furthermore, up to 30% of their fixations are regressions. By fourth or fifth grade, fixation durations and saccade length have stabilized for children as long as the reading material is age appropriate (Rayner, 1986). The rate of regressions continues to decline up through college-age readers (Rayner, Chace, Slattery & Ashby, 2006). Beginning readers, poor readers and dyslexic readers have longer fixations, shorter forward saccades and more regressions than skilled readers (Rayner 1998; Ashby, Rayner & Clifton, 2005).

Fixations tend to be longer in oral reading than in silent reading, ostensibly because the reader has to produce each word as it is read and the eyes, which move faster than the reader can produce words, often stay in place longer so that they do not get too far ahead of the voice (Rayner, 2009). It is well established that the probability to refixate a word during reading of the Roman script depends on where the eyes initially land in the word; the refixation probability is lowest for initial fixations near the word centre but increases progressively as initial fixations deviate to either side of this 'optimal' landing position (O'Regan, 1981; McConkie, Kerr, Reddix, Zola, & Jacobs, 1989; Vitu et al., 1991; Rayner & Fischer, 1996).

Number of fixations

The number of fixations has been used as an indication of numerous phenomena related to reading. Holmqvist, Nystrom et al. (2011. p 413)) have listed the following:

1) Semantic importance. Many researchers agree that the general importance of an object increases the number of fixations in the Area of Interest (AOI). Loftus and Mackworth (1978) found that significantly more fixations were made on semantically informative areas.

2) Search efficiency and difficulty. The number of fixations is believed to be negatively correlated with search efficiency. A low number of fixations could mean that the task is too easy. Conversely, a high number of fixations would be an indication of the difficulty of interpreting information in the task (Ehmke & Wilson, 2007).

3) Experience. Studies have demonstrated that experts make fewer fixations than do novices. This is also true of proficient readers who make significantly fewer fixations than do beginners (Holmqvist, op.cit.).

4) Word properties in reading. Morphological complexity, word frequency and word familiarity also affect the number of fixations with long, unfamiliar and infrequent words receiving more fixations (Clifton, Staub &Rayner, 2007).

5) Dysfunctions. It has frequently been shown that dyslexic readers make more fixations than their non-dyslexic counterparts (Hutzler & Wimmer, 2004).

Perceptual span

Experiments using the moving window measure how large the window of a text must be before readers read normally. Conversely, they also look at how small the window can be before there is disruption in reading. Research using this paradigm has demonstrated that skilled readers of English and other alphabetic writing systems obtain useful information from an asymmetric region extending roughly 3-4 character spaces to the left of the fixation (McConkie & Rayner, 1976) to about 14-15 character spaces to the right of a fixation (McConkie & Rayner, 1975; McConkie & Rayner, 1976 ; Rayner, Well & Pollatsek, 1980 b). The perceptual span for Hebrew readers is asymmetric and larger to the left of the fixation (Pollatsek et al., 1981). Pollatsek et al., (op. cit.) have also found that orthography modulates the size of the span. The span of English is larger than that of Hebrew readers, presumably because English is less "densely packed" than Hebrew and it takes more characters to write the same sentence in English than in Hebrew (Rayner 1994). However, a recent study by Jordan, Almabruk, Gadalla et al. (2014) shows that a leftward asymmetry in the central perceptual span also occurs for Arabic. When readers of English are required to read English text from right to left, the perceptual span also extends farther to the left (Inhoff, Posner & Rayner, 1989) and so it would appear that the perceptual span extends asymmetrically in the direction of reading. Reading skill in general also influences the size of the

perceptual span since beginning readers and dyslexic readers have smaller spans than more skilled readers (Rayner, 1986; Rayner et al. 1989).

Saccades and landing position

Some research has demonstrated that readers use word length and word boundary information in targeting saccades into upcoming words while reading (Plummer & Rayner 2012). Word length information which guides eye movements to the next location, is acquired at about 15 letter spaces to the right of the fixation (Rayner, 1979). Word length is determined by the white space that surrounds each word. "All the brain needs to know is that a word is a letter or group of letters surrounded by a space" (Samuels, Rasinski & Hiebert, 2011, p. 32). Consequently, a white space surrounding a word is an important cue used by the eye in calibrating how far to jump with each saccade. Spaces provide information about an upcoming word's length in parafoveal vision which leads to systematic tendencies in word landing position. When spaces are removed, reading slows down by as much as 30-50 % (Morris, Rayner & Pollatsek, 1990). Removal of space from Romans scripts disrupts both the word identification process and the way the eyes move through the text (McConkie & Rayner, 1975; Pollatsek & Rayner, 1982).

Information about the beginning and ending letters of words, orthographic and abstract letter codes, and phonological information is integrated across saccades (Rayner, 2010). However, there is no strong evidence that semantic or morphological information is integrated across saccades when reading in English. On the other hand, readers of Hebrew do integrate morphological information across saccades (Deutsch, Frost, Pelleg, Pollatsek & Rayner, 2003). The average saccade length is often related to L1. Research on readers of Hebrew (a language which, like Arabic, is written from left to right and also omits short vowels in the written script)

tend to yield shorter saccades of about 5.5 letter spaces than readers of English (Pollatsek, Bolozky, Well & Rayner, 1981).

In terms of landing position, Rayner (1979) demonstrated that readers' eyes tend to land halfway between the middle of a word and the beginning of that word, the preferred viewing location (PVL). The PVL is located slightly to the left of centre for words across all lengths (Plummer 2012). It is generally agreed that readers attempt to target the centre of the words but their saccades tend to fall short. When readers' eyes land at a non-optimal position in a word, they are more likely to refixate that word (Rayner, op.cit.). It has been suggested that the selection of saccade targets is based on information about word boundaries, which is provided by the visually salient space between words (O'Regan, 1981, 1992; O'Regan & Lévy - Schoen, 1987; McConkie et al., 1988; Vitu, 1991; Legge, Klitz, & Tran, 1997; Rayner, 1998, Rayner et al., 1998; Reilly & O'Regan, 1998). This has been evidenced in several studies where it has been shown that removal of space from Roman scripts disrupts both the word identification process and the way the eyes move through the text (McConkie & Rayner, 1975; Pollatsek & Rayner 1982; Rayner, 1986; Rayner et al., 1998. Rayner et al. (1998) demonstrated that without spaces between words, the reading rate decreased by 40–50%, and the PVL moved from near the word centre to the word-beginning.

Word skipping

Just and Carpenter (1980) found that readers fixate an average of 67.8 per cent of the words, with content words being fixated 83 per cent of the time and function words 38 per cent. In this early study, they provided evidence that not every word in a text is fixated, and that the syntactic and semantic components of each word play a role in determining whether fixation occurs. In a more recent study, Schotter, Angele and Rayner (2012) found that extracting information outside

the fovea is an important aspect of information processing which allows for planning of saccade lengths and pre-processing of orthographic information which results in readers skipping 30% of words on average. Rayner (2009) has repeatedly found that content words are fixated about 85% of the time while function words are fixated about 35% of the time. Function words are skipped more because they tend to be short as there is a clear correlation between fixating a word and its length. The likelihood of word skipping dramatically decreases with word length, dropping from about 0.76 for 1- and 2-letter words to about 0.42 for 4-letter words and 0.05 for words of 9–10 letters (Rayner and McConkie, 1976; Vitu & O'Regan, 1995). Words that are 2-3 letters long are fixated approximately 25% of the time, whereas words that are 8 letters or more are almost always fixated. "Furthermore, because the long words extend beyond the word identification span, skipping can occur on the basis of partial information in relation to word identity" (Rayner et al. 2011, p. 514). A study by Fitzsimmons and Dreighe (2011) found that monosyllabic words are skipped slightly more than disyllabic words, indicating that phonological complexity can also influence word skipping. In a meta-analysis of word skipping, Brysbaert et al., (2005) found an average effect size of 5 % for word frequency and 8 % for predictability. Another variable which influences word skipping is the difficulty of the word prior to the target word which is referred to as foveal load. Dreighe (2005) found that there was less skipping under high foveal load conditions.

Regressions

The eyes sometimes return to previous regions of text. Saccadic regressions occur about approximately 10-15 % of the time, in skilled readers (Dussias, 2010; Rayner, 2009; Rayner & Polletsek, 1989). Most are fixations of the word immediately to the left of the last-fixated word (Vitu & McConkie, 2000), but a minority are longer-range regressions to an earlier word or to an

earlier segment of the text. The most obvious explanation of these longer-range regressions (see Rayner, 1998) is that they allow the reader to reread information that they have missed, forgotten, or are unsure about. Regressions are planned and executed differently from forward-directed saccades (McConkie, Kerr, Reddix, & Zola, 1988; Vitu & McConkie, 2000). Readers make more regressions when the text is complex (Rayner & Pollatsek, 1989), when the topic changes (Hyönä, 1995), when the text contains grammatical errors or ambiguities (Inhoff, Greenberg, Solomon & Wang, 2009), and when spelling errors are introduced into the text (Zola 1984; Frazier & Rayner, 1982). All of this would suggest that readers make regressions to reread or check previously read words. Some studies have found that readers with dyslexia execute more regressions (Lefton, Nagle, Johnson & Fisher (1979) as do Alzheimer's patients (Lueck, Crawford, Hansen & Kennard, 2000).Throughout early school years, the number of regressions decreases up to university. However, regressions begin to increase again in old age (Rayner, Reichle, Stroud, Williams & Polletsek, 2006).

Table 2.1 below gives a short definition of each metric used in my research with a brief explanation of how each metric differentiates skilled versus less skilled readers (for a more in-depth description, see Chapter 4.2).

Table 2.1

Explanation of eye movement measures and how they differentiate skilled from less skilled readers

Measures & Definitions	**Skilled Reader vs Less Skilled Reader**	**Sources which have informed the hypotheses**
Fixation The period of time during which the eye remains relatively stationary and reflects when information is being encoded.		
Number of Fixations	As ability increases, the number of fixations decreases.	Holmqvist, 2011; Rayner, 1998
Fixation Duration	As ability increases, the duration of fixations decreases.	Holmqvist, 2011; Rayner 1998
Saccade A high velocity jump executed by the eye to bring information into foveal vision.		
Number of Forward Saccades	As ability increases, the number of saccades decreases.	Hyönä & Olson, 1995; Rayner, 1998; Rayner et al., 2006
Length of Forward Saccades	As ability increases, the length of saccades increases.	Rayner, 1998
Regressions "In order for a saccade to be a regression, the saccade needs to move in the opposite direction to the text but not necessarily in the opposite direction to the previous saccade". Holmqvist, 2011, p. 263		
Number of Regressions	As ability increases, the number of regressions decreases.	Ashby, Rayner & Clifton, 2005
Length of Regressions	As ability increases, the length of regressions increases.	Rayner, 2019; Dussias, 2010

Despite the promising contributions eye tracking research might bring to reading comprehension, Rayner, Chace, Slattery and Ashby (2006) are hesitant to advocate its use as a diagnostic tool. They have noted that reading is a learned skill and there is large variability between individuals. They suggest that eye movement data could be used in conjunction with reading comprehension assessments to formulate individual education plans. They also point out that effect

sizes are in the order of fractions of a second and so the logistics of obtaining reliable data from one reader is onerous.

2.3. Conclusion

In Chapter 2, I discussed the cognitive processes involved in reading and the differences between reading in the L1 and reading in the L2, with specific attention to the Arabic L1 learner of English. These particular learners tend to experience great difficulty in learning to read in English due, in part, to the nature of the Arabic language and its linguistic distance from English. I have given a brief history of eye movement studies in investigating the reading process. I have also briefly reviewed selected aspects of eye tracking methodology and have intimated the potential contributions that eye tracking research might bring to reading research.

As Hayes-Harb (2006) noted, understanding the specific word recognition problems Arabic L1 learners of English experience has the potential to help teachers develop effective strategies for teaching these learners to read in English. In addition, if the learners are consciously aware of their word identification difficulties, it is possible that conscious strategies may help them to process English words more efficiently. Study One is unique in that it is the first study to compare the eye movements of Arabic L1 EFL learners and skilled English L1 readers. Contrasting native and non-native reading patterns can potentially generate information as to how different the learner's reading behaviour is to that of a typical skilled reader. This knowledge can then inform pedagogical interventions to help remediate reading difficulties.

Arabic L1 learners of English do not appear to experience optimum success with mainstream reading materials despite the claims by writers and publishers that these textbooks can be used in any geographical location with learners of various language backgrounds. In Chapter 3,

I will describe specific reading interventions that I hypothesize will be effective in the reading instruction of Arabic L1 EFL students. I will then discuss the concept of 'raising awareness' in the foreign language classroom and how eye tracking technology has been used to measure it.

Chapter 3: Reading interventions and awareness raising

Introduction

In Chapter 3, I will review the five types of reading interventions that I implemented in Study Two: textual enhancement, phonological awareness training, training in word recognition and automaticity, training in oral text reading and spelling instruction. I will then discuss the Noticing Hypothesis and measurement of the attentional component of Noticing.

3.1 Reading interventions and instruction

In its comprehensive report, The National Reading Panel in the US (National Reading Panel, 2000) recognized reading fluency as one of the five essential elements to be considered in reading instruction, along with alphabetics, comprehension, teacher education, and computer technology. This report was the impetus behind the now burgeoning implementation of reading intervention programmes. There are different types of reading interventions such as computer assisted instruction, phonics instruction, fluency practice, use of text comprehension strategies and numerous others. In Study 2, I implemented five interventions, the choice of which was based upon current literature and my fifteen years' experience teaching Arabic L1 learners of English.

3.1.1 Textual enhancement

Sharwood Smith (1993) proposed that through enhancement, the teacher manipulates the form in such a way that it becomes more perceptually salient to the L2 learner than it normally would be. It increases the likelihood that the learner will notice the form, which is the first step toward learning it. The underlying assumption is that paying attention to forms in the input is a necessary pre-condition for learning them (Leow, 1997; Robinson, 1995: Schmidt, 1990, 2001;

Sharwood Smith, 1991). However, Sharwood Smith (op.cit) also noted that although the input is 'engineered' by the teacher to increase the saliency of a form, this is no guarantee that learners will notice it. He posed the question, "Why is it that L2 learners typically appear to ignore a vast mass of evidence and continue, obstinately, to operate with a system that is in contradiction with the target norms as manifest in the input?" (Sharwood Smith, 1993, p. 68). Han, Parks and Combs (2008) presented three plausible reasons for this phenomenon. First, learners may lack grammatical sensitivity to the features of the target language input. Secondly, there are certain features in the input which are inherently non-salient and hence their presence often escapes the learners' attention. Lastly, the learners' L1 may hinder their ability to notice certain linguistic features in the input. Therefore, the failure to benefit from input may arise from a combination of lack of noticing ability and poor input characteristics, such as lack of perceptual salience or noticeability (Schmidt, 1990). Different external manipulations have been used to increase the salience of input, including manipulation of frequency (input flooding), visual salience (typographical or textual manipulation) and corrective feedback such as repetitions or recasts.

Textual enhancement (TE) is an implicit focus-on-form technique, through which learners' attention is drawn to a language form during an otherwise meaning-focused interaction. TE uses visual enhancement methods such as underlining, bold facing, colour-coding, italicizing, capitalizing or using different fonts as a means to promote the processing of linguistic items. The perceptual salience created by highlighting the input is intended to draw the learners' attention to the form and, once the first step is successful, learning of the attended form is expected to occur, based on the premise that attention is what mediates input and intake (Izumi, 2002). TE is considered an 'implicit and unobtrusive' way of directing learners' attention to targeted forms

(Nassaji & Fotos, 2011) and is more explicit than input flooding (an artificially increased number of target forms) but less explicit than rule explanation (White, 1998).

There is a general consensus that paying focal attention to items in the input is a necessary condition for its sustained processing (Leow, 1997; Schmidt, 2001; Sharwood Smith, 1993; Tomlin & Villa, 1994). However, some studies have found that a competitive relationship exists between learners' capacity to attend to form and their ability to attend to meaning (Lee, 2007). Lee (op.cit.) demonstrated that participants were better able to correct written sentences which contained incorrect passive forms after exposure to texts flooded with enhanced passive forms. However, they scored lower on comprehension measures. In a replication study, Winke (2013) found that enhancement did not significantly increase gain scores on form correction, nor did it detract from comprehension. Using eye tracking, she found that enhancement significantly impacted noticing of the passive forms through longer gaze durations and rereading times. She noted that it promoted noticing but without further explicit instruction, it appeared to have done little else (op.cit, p. 324). Jahan and Kormos (2014) also found that exposure to textually enhanced input facilitated the development of form-function mapping of "be going to" and "will", but that the enhancement alone was not effective in helping the learners to fully understand the targeted meanings. Simard and Foucambert (2013) demonstrated that language learners fixate significantly longer on enhanced text than on unenhanced text and they proposed that enhancing linguistic forms "is a good way to increase salience of the targeted language features" (p. 18). In their 2016 study, Indrarathne and Kormos investigated how attentional processing of the grammatical construction *causative had* varied under four different input conditions. Contrary to the findings of Simard and Foucambert (op.cit.), they found that the attentional processing of the participants in the enhanced condition did not seem to differ significantly from that of the participants in the unenhanced

condition. They suggested that a possible reason for the discrepancy in findings might be the nature of the enhancement itself. Simard and Foucambert (op.cit) enhanced the text by underlining while Indrarathne and Kormos (op.cit) enhanced their texts with bold. They concluded that bolding as a mode of enhancement was not successful in generating additional attentional processing.

In a recent eye tracking study, relevant to my research, Alsadoon and Heift (2015) explored the impact of textual input enhancement on the noticing and intake of vowels by 30 female Saudi Arabian students studying in various ESL schools in Canada. Participants were randomly divided into an experimental and control group. The treatment consisted of a reading task with and without textual input enhancement of the target word and its vowel(s). Eye tracking data were collected for all participants during the treatment phase and an immediate post-test was administered directly after the treatment, with a delayed post-test administered 4 weeks after the treatment. Intake of the targeted forms was measured by a multiple-choice recognition task using 36 sentences. Participants were asked to select two words from two separate drop-down menus, each of which provided 3 word choices. The first drop-down menu tested the orthographic vowel knowledge of the item by presenting a correct English word and two variations which contained the same consonant structure but different vowels (wanter,winter,wenter). The second drop-down menu measured the participants' knowledge of word meaning by displaying three Arabic words, one of which was the correct translation of the test item. From an initial database of 264 common 'vowel blindness' errors (Bowen, 2011; Ryan& Meara, 1991; Saigh & Schmitt, 2012), 36 test items were randomly chosen. The treatment consisted of 36 trials for both groups of participants. Each trial consisted of three separate display screens. In the experimental condition, the first screen displayed a sentence in which the target and its vowel(s) were textually enhanced with three typographical cues; the target word was underlined and the vowels were bolded and in red colour. In the control

condition, there was no textual enhancement. The second screen prompted the participants to choose the correct meaning of the target word in a multiple-choice task. The third screen provided feedback to the participants on their choice of word meaning. Eye tracking data were recorded for First Fixation Duration, Refixation Duration, Rereading Duration and Total Duration.

The researchers attempted to answer the following research questions: 1) Does textual enhancement have a significant effect on Arabic ESL learners' intake of vowels in a target word, as measured by a multiple-choice word recognition task? 2) Are there significant effects of textual input enhancement on Arabic ESL learners' intake of vowels in a target word over time, as measured by a multiple-choice task in an immediate and delayed post-test? 3) Does textual enhancement significantly draw the visual attention of Arabic L2 learners of English to a target word and its vowels? 4) Do longer eye fixations on a target word and its vowel(s) result in a significant increase of orthographic vowel knowledge?

Alsadoon and Heift (2015) reported that results for RQ 1 indicate that the experimental group had fewer word form errors; the two groups did not differ in the frequency of errors in word meaning. Results for RQ 2 showed that while both groups made fewer mistakes with word forms on the delayed post-test than on the pre-test, the difference was "more pronounced" (p. 67) with the experimental group. With regards to RQ 3 and 4, the authors reported that on average the participants fixated about 13 times longer on the target form than the control group. Refixation showed that the experimental group spent more time on the target form than the control group. Rereading Duration indicated that the experimental group fixated on the target word longer than the control group. The authors suggested that the textual enhancement drew the experimental group's visual attention to the target forms and their vowels by rereading the target forms and focusing on them longer than the control group. In summary, they reported that the experimental

group significantly outperformed the control group on the word form of the test items, and there was a significant gain in word form from the pre-test to both the immediate and delayed-post tests for the experimental group, "implying that the treatment significantly improved the learners' orthographic vowel knowledge and thus reduced their vowel blindness" (p. 69).

There are several limitations to this study, in terms of methodology, data analysis and interpretation of findings. No information was included concerning the algorithms used to detect the fixation and saccade locations. As Salvucci and Goldberg (2000) pointed out, researchers often have minimal information to guide them in their decision of which algorithm to use in a given situation. This often "leads to often haphazard application of different algorithms making it difficult to compare results derived by different methods of identification" (p. 72). In addition, it is not clear why the four particular metrics were chosen. The authors state only that the "four main variables are widely used because they accurately capture the attentional and cognitive processes that take place during the reading process of a target word" (p. 64). In addition, averages were calculated between the First Fixation and other fixations. In reading research, the first fixation on a word appears to be associated with lexical activation, and later fixations with integrative processes. Inhoff and Radach (1998) point out that it may be confounding to form averages over fixations that are qualitatively different. Additionally, it appears that the Areas of Interest (AOIs) were drawn on individual words and not on vowels or consonants. Therefore, research questions 3 and 4 cannot be fully investigated, as it is not possible to ascertain from the AOIs if participants looked at the vowels within the target words.

In terms of the participant pool, it seems questionable that 35 participants of varying levels of proficiency (from beginner to advanced) could be randomly assigned to two different groups where a pre-test showed there were no significant differences between the groups. An additional

problem is that no information is provided as to the attrition rate i.e. the number of participants who were eliminated due to participant-specific tracking difficulties or low quality data. Alsadoon and Heift (2015) claim their results show that longer fixation durations resulted in a significant increase in vowel knowledge, textual enhancement drew the experimental group's attention to the target words and their vowels, and that the textual enhancement was effective in the training of vowel blindness. However, the nature of their data does not support these claims because their AOIs were drawn around words and not vowels.

Generally speaking, lack of congruence in the findings of the studies on input enhancement is a natural consequence of the numerous "methodological idiosyncrasies" characterizing the individual studies (Lee 2007). Jahan and Kormos (2014) proposed several plausible explanations for the differing results of these studies. First, textually enhanced forms may not always be noticed by learners because salience created externally by teachers may not correspond with learners' internally generated salience. Second, prior knowledge of the construction being studied may facilitate noticing of the textually enhanced conditions. Thirdly, most of the researchers utilized short-term treatment with somewhat limited exposure and only two studies, Izumi (2002) and Simard (2009), included both experimental and control groups. In their 2008 meta-analysis of input enhancement and grammar learning, Lee and Huang pointed out additional explanations for the discrepancy in findings. The first relates to the participants' proficiency. Generally speaking, researchers rely on the participants' institutional status to classify their proficiency levels. They suggest that as learner proficiency could potentially be a factor affecting the ability to focus on forms, additional measures to assess proficiency should be incorporated in future studies. Finally, they noted a possible publication bias in the studies they investigated. Publication bias suggests that studies which show no effect are often not published. Therefore, the studies actually published

in refereed journals might not be representative of the true effects of visual enhancement across different educational contexts.

Still other researchers proposed that enhancement might not be enough in itself, and other supplementary instructional elements should be added to this technique (Indrarathne & Kormos, 2016; Izumi, 2002; White, 1998). Spada and Tomita (2010) in a meta-analysis of 41 studies investigated the effects of explicit and implicit instruction on the acquisition of grammatical features and found that effect sizes were larger for explicit over implicit instructions for both simple and complex grammatical features. Winke (2013) pointed out that learning is not instantaneous. She also noted that learners were not told to pay attention to the forms as would happen in a real classroom and therefore outcomes from studies that investigate enhancement may not be generalizable to the classroom because of the artificial nature of these studies. She indicated that future input enhancement research be conducted in conditions more representative of how teachers use input enhancement in the real world (op.cit, p. 343). Godfroid and Winke (2014) also called for more research to understand how enhancement and other implicit form-learning conditions affect the allocation of attention and L2 development.

3.1.2 Phonological Awareness Training

Another possible way of enhancing students' word-level reading skills is by phonological awareness training. Phonological Awareness (PA) is the ability to recognize that words are made up of a variety of sound units. As children develop phonological awareness, they come to understand that words are made up of small sound units (phonemes). They also learn that words can be segmented into larger sound "chunks" known as syllables and each syllable begins with a sound (onset) and ends with another sound (rime). Phonemic Awareness also involves an

understanding of the ways that sounds function in words but it deals with only one aspect of sound: the phoneme. PA is a powerful predictor of reading success (Ehri, 1984; National Reading Panel, 2000). A large quantity of research has demonstrated the beneficial effects of PA instruction on reading for English first language children (Ehri et al., 2001; Snow, Burns, & Griffin, 1999). Research has indicated that PA instruction has beneficial effects for ESL and EFL learners as well (Yeung, Siegel & Chan, 2013). The National Reading Panel (NRP) (2000) in its meta-analysis examined whether PA instruction was significantly more effective than alternative forms of training in helping children acquire PA (specifically, phonemic awareness) and enabling them to apply this skill in their reading. Results showed that PA instruction is beneficial under different teaching conditions with a variety of learners. The findings also indicated that teaching children to manipulate the sounds in a language helps them learn to read. In addition, the effects of the PA training on reading lasted well beyond the end of training.

Lesaux and Siegel (2003) conducted a longitudinal study in 30 schools in a Canadian school district. 972 English L1 students and 188 ESL students participated in the study from kindergarten to Grade 2. The ESL students were from a wide variety of language backgrounds: Cantonese, Mandarin, Korean, Spanish, Farsi and Polish. PA instruction consisted of classroom-based small group activities. The ESL children had difficulties in kindergarten and performed more poorly on tasks of rhyme detection, pseudo word repetition, memory for sentences, syntactic awareness and rapid naming. However, by grade 2, the ESL children had acquired the sound-symbol relationships to the extent that they were both reading and spelling as well as, and in some cases better than, the English L1 children. Specifically, the performance of the ESL children was significantly better than that of the L1 children on tasks of word reading, rapid naming, real and

non-word spelling and even arithmetic. Lesaux and Siegel (op.cit.) concluded that kindergarten PA instruction is "as effective for ESL children as for English L1 children" (p. 1016).

In a more recent study, Yeung, Siegel and Chan (2013) examined whether PA instruction over a 12-week program helps to promote phonological awareness and reading skills of kindergarten Chinese learners of English in Hong Kong. 76 children from three kindergartens completed the instructional program: 38 children in the instructional group and 38 in the comparison group. There were three major findings: first, the PA instruction was found to facilitate the acquisition of phonological awareness (at syllable, rhyme and phoneme levels), expressive vocabulary, word reading and also word spelling, better than the comparison instruction which consisted of vocabulary learning and writing tasks but no direct instruction in phonological awareness skills. Second, changes in phonological awareness predicted improvements in word reading and spelling after controlling for the effects of general intelligence, oral language skills and the initial ability of the children participating in the program. Finally, phoneme awareness was demonstrated as the most important unit of phonological awareness in explaining beginning L2 reading of Chinese L1 EFL children (p. 697).

3.1.3 Training in word recognition and automaticity

Training in word recognition and automaticity is another potential method to help develop word-level reading skills. Both the Automaticity Theory (DeKeyser, 2001; LaBerge & Samuels, 1974; Rasinski & Samuels, 2011; Segalowitz, 2003; Segalowitz & Segalowitz, 1993) and the Verbal Efficiency Theory (Perfetti, 1985, 1988; Perfetti & Lesgold 1977) proposed that attention and working memory are limited in capacity. In text comprehension, if lower-level processes such as word meaning retrieval do not take place 'automatically', they may require attentional capacity,

to the detriment of the higher-level comprehension tasks. Segalowitz and Segalowitz (1993) noted that L2 word recognition fluency involves two qualitatively different changes: simple speedup and automatization. They proposed that in the initial stages of the development of L2 word recognition skills, there is merely a speedup of performance. This simple speedup stage is qualitatively different from the subsequent automatization stage because automatization is associated with restructuring word recognition mechanisms or with increasing cognitive efficiency (Segalowitz & Segalowitz, 1993, 1998).

Fukkink, Hulstijn and Simis (2005) posed the trenchant question "Would it not make sense to train second language learners in methods of quick retrieval of word meanings after they have been exposed to these words the first time?" (p. 54). The authors conducted two classroom-based studies which investigated automatization of lexical access with Dutch L1 EFL students. In experiment 1, during two 40-minute class periods within a single week, participants worked with the help of laptops to complete two types of exercises: a translation task and a cloze task. Experimental comparisons involved the performance differences between trained words, context words and control words. Before and after the training sessions, the students were given a word recognition test (WRT) which consisted of 100 words and 90 pseudowords. The 100 target words were randomly subdivided into 40 words for training, 40 control words, which only appeared in the pre-and post-training WRT, and 20 context words, which occurred in the carrier sentences of the second exercise type of the training (carrier sentences are those which are used to present test words; they 'carry' the test words). The context words were not the target of the training, but, in contrast to the control words, they did appear in the exercises at least once. Responses were coded for accuracy and reaction time. The training was not limited to word recognition but included practice in word meanings. Results suggested that both a quantitative speeding up and a qualitative

automaticity process occurred for the recognition of the trained words. In experiment 2, training was aimed at accelerating and automatizing lexical access of familiar and unfamiliar words. Students were randomly assigned to each of four classes to either condition A or B. A different set of target words was trained in four training sessions in each condition, which they labelled A words and B words. Before and after training, participants completed a word recognition test containing word stimuli from sets A and B. The study entailed seven regular EFL lessons. Training consisted of four exercise types: cloze, translation, column and square format. Words were presented in sentences for the cloze and translation, while the square[2] and column[3] formats provided practice using collocations, prepositional phrases and phrasal verbs.

After the training, they completed a test of comprehension consisting of subtests A and B. The authors explained that this design allowed them to test the hypotheses twice i.e. they could measure the effect of the training of A words on participant performance in the A test of reading where participants from the A condition served as the experimental group and students from B served as the control. Conversely, they assessed the effect of training in B words on performance on the B test of reading comprehension where the experimental group consisted of B participants and the control consisted of A participants. Results showed that training enhanced participants' lexical access in the L2 although the actual progress due to the training was greater for reaction time than for a qualitative change of lexical access. The researchers interpreted the findings from both experiments as lending support more for an acceleration interpretation than for a qualitative change interpretation of automatization. They concluded that "automatic lexical access helps in

[2] In the column format, the target word appeared on the left of the screen in a short unfinished sentence prompt. Three grammatical, but only one semantically correct, distractors appeared on the right.

[3] In the square format, a word was presented in the middle of the screen. Eight words or phrases appeared in a clockwork arrangement around the target word. Participants had to indicate whether the combination of the target word with the words or phrases surrounding it was semantically possible.

processing the words in a text but this is only one of the many sub-processes operating in the complex nature of L2 reading comprehension" (p. 72). The researchers tested for, but did not find, a significant improvement in reading comprehension after 2 days of word recognition training. However, reading fluency is achieved through the development of automaticity and requires extended periods of implicit learning. A large recognition vocabulary size is also necessary for fluent reading performance. This skill set is only learned gradually and is not always easy to detect in shorter training studies (Grabe 2010).

In his 2008 study, Akamatsu provided word recognition training as part of a regular reading class with first year university Japanese L1 EFL students. The participants were given 7 word recognition training sessions, one session per week. In each session, they carried out word chain tasks in which they had to draw lines between words written with no spaces (for example sunbendgivebearpen) as fast and as accurately as they could. Results showed that the learners benefitted from word-recognition training in speed and accuracy. They recognized the target words more quickly and more accurately than they did before the training. Akamatsu suggests that training in EFL word recognition improves word recognition in speed and accuracy.

3.1.4 Training in oral text reading fluency

A further means of improving students' word-reading skills is training in oral text reading. Fluent reading is what most good readers habitually do when they read a variety of texts, especially in their L1 (Grabe, 2012). Fluency is the ability to read a text quickly, accurately, and with proper expression. It combines accuracy, automaticity and oral reading prosody and is a factor in both oral and silent reading which can limit and support comprehension (Grabe, 2009; Kuhn, Schwanenflugel & Meisinger, 2010; Rasinski & Samuels, 2011). Reading fluency has recently

become the focus of numerous studies (see, for example, Benjamin & Schwanenflugel, 2010; Binder, Tighe, Jiang, Kaftanskim, Qi & Ardoin, 2013; Jiang, Sawaki, & Sabatini 2012; Klauda & Guthrie, 2008; Veenendaal, Groen & Verhoeven, 2015; Yamashita & Ichigawa, 2010).

The National Reading Panel (2000) concluded that guided oral reading procedures had a significant and positive impact on word recognition, fluency and comprehension across a range of grade levels. Despite it being an essential component of reading fluency (Kuhn & Stahl, 2003; Rasinski & Hoffman, 2003), prosody is often overlooked in studies on fluency and reading fluency instruction (Allington, 1983; Rasinski, 2006). Prosody refers to reading with expression; it is sometimes referred to as the melodic element in reading. Fluent reading is usually characterized by readers who read at an appropriate rate, but who also convey meaning through their voice using pitch, stress, and appropriate phrasing (Dowhower, 1991; Schreiber, 1991). Stahl and Kuhn (2002) define prosody as the "ability to make oral reading sound like spoken language" (p. 582). Proficient prosodic readers divide text into meaningful units marked by such prosodic cues as pauses, varied duration of pauses, the raising and lowering of pitch, and lengthening of certain vowel sounds (Binder et al., 2008). Prosody comprises a series of features including pitch or intonation, stress or loudness, and duration or timing, all of which contribute to an expressive rendering of a text (Allington, 1983; Dowhower, 1991; Schreiber, 1991). Additionally, prosodic reading includes appropriately chunking groups of words into phrases or meaningful units in accordance with the syntactic structure of the text. Prosody is measured using one of two approaches: subjective rating scales and spectrographic measures. For obvious reasons, subjective rating scales are most often used for evaluation in classroom settings. Klauda and Guthrie (2008) developed a fluency rubric which assesses five dimensions of prosodic reading: passage expressiveness, phrasing, pace, smoothness, and word expressiveness. Studies using these scales have consistently found that

prosody is related to reading comprehension (Cowie, Douglas-Cowie, & Wichmann, 2002) and that better prosody is typically observed in students with greater reading achievement). Better readers pause less frequently while reading, decrease their pitch at the end of sentences, and do not always stress words as heavily as do many poor readers (Klauda & Guthrie, op.cit).

Jiang et al. (2012) investigated the relationship between oral reading fluency and reading comprehension with 200 adult Chinese L1 EFL learners. The results showed that oral passage reading fluency correlated significantly with comprehension. In an American study by Rasinski, Rikli and Johnston (2009) with English L1 students in grades 3, 5 and 7, reading specialists and teachers were trained to listen to electronic recordings of students' oral reading and assign a score for each of the prosodic dimensions listed in the Multi-Dimensional Fluency Scoring Guide (Zutell & Rasinski, 1991). One of the findings was that fluency continued to be highly correlated with silent reading comprehension at grades 5 and 7, thus suggesting that fluency continues to have importance beyond the primary grades and middle grades.

Binder et al. (2013) investigated the relationship between prosody, and reading comprehension in English L1 adults with low literacy skills compared to skilled readers. The participants included 57 adults from Adult Basic Education classes with a wide range of ages and diverse ethnic backgrounds. Data were also collected from 28 female college students who served as the skilled reader comparison group. Participants read a narrative passage orally, and information was extracted from the recordings on pauses and pitch changes using computer software. Results showed that adults with low literacy skills paused longer than skilled readers. They also paused at a larger number of punctuation marks than the skilled participants. They stumbled on more words (i.e. word intrusions) and had more irrelevant pauses within sentences (i.e. sentence intrusions). In addition, their readings of questions lacked a change in pitch. The

authors concluded that decoding and word recognition skills were related to pauses while reading. In other words, readers with lower skills made longer and more frequent and inappropriate pauses. As an explanation for the large number of pauses, the researchers claimed that readers will compensate for poor decoding and/or working memory skills by slowing the rate at which they read, pausing more often while reading, and re-reading prior text because such pausing serves the purpose of providing more time for cognitive processing.

3.1.5 Spelling intervention and instruction

To account for how word recognition skills develop, some researchers (Ehri, 2005; Ehri & Snowling, 2004; Perfetti, 1992, 1997) claimed that word recognition skills improve as the quality of the spelling knowledge in the orthographic lexicon develops. Perfetti (1997) defined spelling as a "linguistic skill that involves encoding linguistic forms into written forms – the linguistic units, phonological strings, morphemes and words – are provided by the language" (p. 22). Spelling is a language skill supported by several linguistic knowledge sources, including phonemic, orthographic and morphological knowledge (Masterson & Apel, 2010). When learners write a word, they can summon one or more sources of linguistic knowledge to spell that word. They may use their knowledge of speech sounds, or phonemic awareness, to identify the sounds in a word. Then, when the sounds are identified, they can access their orthographic knowledge to 'translate' the speech into writing. This can be done in two ways. When learners have well-established memories of specific written words or mental graphemic representations (MGRs), they can directly access and represent these word specific spellings in their writing. When MGR has not been established, spellers can use their orthographic knowledge to represent the sounds with one or more graphemes or orthographic patterns. Current theories of spelling development suggest that

these different sources of knowledge are accessible and used by young spellers early in their spelling development (Masterson & Apel, 2007, Siegler, 1996a).

Contrary to the view of spelling as a rote memorization skill, spelling requires active consideration of the sounds, patterns and meaning of written language (Masterson & Apel, 2000; Moats, 2000). Rote approaches provide little or no instruction for the development of the linguistic sources of knowledge that support spelling. However, research has demonstrated significant improvement in spelling when one or more of these underlying linguistic sources of knowledge are taught (Apel & Masterson, 2001; Berninger et al., 2003; Kelman & Apel, 2004). Compared to other means of teaching spelling, a strategic spelling approach typically leads to better results (Wanzek et al., 2006). A multilinguistic spelling approach focuses on teaching students when and how to use their linguistic awareness skills to spell words i.e. to provide practice of their phonemic and orthographic awareness skills and activities that teach the strategy of segmenting words into their individual phonemes and then linking each sound to a letter.

Ehri (1997) and Perfetti (1997) observed that spelling and reading appear to be "both sides of the same coin", even though spelling is more challenging for most students to learn. Incomplete or inaccurate spelling representations or knowledge will result in less efficient and in some cases, less accurate word recognition skills (Ehri op.cit.; Perfetti, 1992). Ehri (1987, 1989) proposed that spelling contributes to reading development by shaping knowledge of phonemic awareness, strengthening the student's understanding of the alphabetic principle. When a student encounters a new spelling, his examination of how the letters and phonemes in the word are arranged, gives additional cues about the phonemic structure of the word, thus making sight words easier to remember. One source of evidence supporting the view that spelling and reading are closely linked is the finding that children who are good readers are usually good spellers (Ehri, 1987).

Ehri (1980) had second graders practice reading eight identically pronounced spellings of made-up words naming animals (e.g. weeple while others practiced wheople). At a later time, the students were asked to write the words from memory. She found that if they had read WH, they always wrote WH, never just W. She believed her findings showed that students did not just decode print to speech, store pronunciations in memory, and later invent spellings of the words. Rather, they stored the specific letters for those words in memory. Ehri and Wilce (1979) found that students remembered the spoken words better when they had seen the spellings than when they had not. Ehri and Rosenthal (2007) conducted two laboratory experiments to examine whether spelling improves the learning of new vocabulary. Their hypothesis was that learners would learn the pronunciations and meanings of new words better when they saw the spelling of the words during study period than when they did not. Spelling helped both second and fifth graders, indicating that the effect of spelling is not limited developmentally to the period of beginning to read. Their explanation was that "grapheme-phoneme connections would be activated by spellings and as a result should secure the pronunciations and meanings of words in memory earlier during learning" (p. 397).

Spelling of words is especially valuable for students who are learning English as a Second or Foreign Language. Spelling helps to clarify phonemes in pronunciations. When students listen to spoken words while inspecting their spelling, more precise representations are secured in memory. Unfortunately, very little research has been conducted on the effects of ESL/EFL spelling interventions, although research with young children has shown that English spelling knowledge and English word-reading skills are closely related (Chiappe et al., 2007; Geva & Zadeh, 2006; Wade-Woolley& Siegel, 1997).

Graham, Harris and Chorzempa (2002) examined the contribution of supplemental spelling instruction to spelling, writing, and reading among EFL students in Grenada. Second grade children experiencing difficulties learning to spell participated in 49 twenty-minute sessions designed to improve their spelling skills. In the intervention programme, the experimental group was provided with 'systematic, explicit spelling instruction' for an average of 15-20 minutes twice a week. The control group followed ordinary curricular materials. A delayed post-test was administered 6 months later. In comparison to their peers in the control group receiving mathematics instruction, students in the spelling condition made greater improvements on norm-referenced spelling measures, a writing-fluency test, and a reading word-attack measure.

Anastasiou and Griva (2012) used a Morphological Processing Spelling Approach (MPSA) in a Grade 6 EFL class in Greece. MPSA is a type of morphological processing strategies training. It provides explicit and systematic instruction in word-level spelling where students do dictation of a meaningful text. Seven dictation tasks, each one including a specific morphemic pattern recycled in ten different words were used during 45-minute sessions in two terms of the school year. Twenty-five students participated in the control group while 23 were in the experimental group. Pre- and post-tests of spelling were administered to all participants. The post-tests revealed that the students in the treatment group scored significantly higher than the control group in spelling tests.

Fender (2008) found that most of the Arabic L1 learners of English in his study demonstrated spelling knowledge of within-word spelling patterns with long and short vowels (e.g. train, reach catch, dress) but exhibited more difficulty with multisyllabic words that included spelling patterns across syllables (e.g. decision, knowledge, responsible). He suggested interventions such as being shown how to segment words into syllables and "acquiring an

understanding of syllable-level-spelling patterns with open and closed syllables." (p. 17). This understanding may be crucial in helping learners acquire how English spelling patterns at the phoneme, syllable, morpheme or word levels correspond to pronunciations. Then when readers see and pronounce a word, an enriched awareness of spelling-sound relationship can be utilized, which then may potentially secure a word's spelling in memory.

These findings are important for teachers as they need to recognize the contribution that exposure to spelling can make and consequently include spelling instruction in their teaching. Many teachers believe that reading is the primary way that students acquire new vocabulary. However, it is not uncommon that when students read independently they skip words they do not know and fill them in by guessing (Goodman 1970). Unfortunately, this guessing strategy does little for vocabulary learning. Perfetti, Rieben and Fayal, (1997) noted that reading by itself will not dramatically improve spelling because reading does not practice the full orthographic process demanded by spelling. Moreover, it is spelling itself that is most effective at improving the quality of the word representation. "Practice at spelling should help reading more than practice at reading helps spelling" (p. 30).

3.2 Noticing and awareness

The hypothesis that input does not become intake for language learning unless it is noticed, was first proposed by Schmidt (1990, 1995, 2001, 2010). Schmidt claimed that if L2 learners were to acquire any aspect of the second language, they would need to notice the relevant material in the linguistic environment. Noticing, as he originally defined it, was the brain registering something new, even if the learners did not understand how the new information worked or even if they could not remember the noticing 'event' at a later time. His critics pointed out that it is

difficult to distinguish absence of noticing from inability to remember and report the noticing at a later time. Subsequently, Schmidt (1995, 2001) weakened his claim to conclude that the more L2 learners notice, the more they learn and that learning without noticing, even if it exists in other domains of learning, plays a minimal role in language learning.

Schmidt's Noticing Hypothesis (NH) was the first psychologically oriented theory that succeeded in achieving mainstream status in SLA theory because it addressed two fundamental issues: the role of explicit, conscious learning and the explanation of why only a selected portion of input becomes intake during the learning process (Dörnyei, 2009). The core of the NH is that only those parts of the input that the learner notices become available for intake and effective processing, and to learn some new information effectively, the learner needs to attend to it at the first encounter. From a theoretical perspective, Schmidt (1995) defined two levels of awareness: 1) noticing which entails attention with subjective awareness and 2) understanding which entails the ability to analyse, compare or test hypotheses. He proposed that noticing is necessary for second language acquisition and that understanding is facilitative but not required.

Ample evidence has been provided that noticing with awareness, and even more so with understanding, is facilitative of L2 learning. Using a crossword puzzle to manipulate the learners' attention when exposed to examples of Spanish stem-changing verbs, Leow (1997, 2000) found that those participants who showed a higher level of awareness (understanding) learned the most; those who noticed instances but attempted no generalizations learned next to most; and those who reported no noticing experienced no learning. Mackey (2006) found that learners reported more noticing when feedback was provided, and learners who exhibited more noticing developed more in their production of question formation than those who exhibited less noticing.

However, Williams (2005) provided empirical evidence of unaware learning. Awareness was assessed by post-instruction interviews and he concluded that at least for some individuals, it is possible to learn form-meaning connections without awareness of what those connections are. Hama and Leow (2010) replicated Williams' (2005) study in which awareness was assessed at the encoding/training stage. Their research made no methodological changes from Williams' study but even when using the same classification system, they failed to find above chance performance for unaware learners on generalization items. Godfroid, Boers and Housen (2013) found that learners spent more time processing unknown pseudo words than their matched controls. The longer participants looked at a pseudo word, the more likely they were to recognize it later, giving more support to the notion "that attention is crucial for encoding in memory" (p. 21).

Noticing is "a hybrid concept because it entails both attention and awareness" (Godfroid, Boers & Housen, 2013, p. 485). Indrarathne and Kormos (2016) observed that there are two unsettled issues in exploring noticing: the first is "the role and degree of awareness involved in noticing and the other the establishment of a critical threshold of processing above which one can claim that noticing took place" p. 7). It is important for researchers to be explicit in terms of which process (attention or awareness) their measures purport to assess. The focus of my two studies will be on the attentional constituent of noticing using the definition that Godfroid et al. (2013) propose based upon Lamme (2003) who stated that attention is a "selection process where some inputs are processed faster, better or deeper than others so they have a better chance of being memorized" (p. 14). Based upon the operational definition provided by Godfroid (op.cit), I will operationalize attention as a quantitative variable which is reflected in a participant's eye fixations and saccades during reading.

3.2.1 Measuring the attentional component of noticing

The question of whether there can be learning without awareness continues to be debated within both psychology and applied linguistics (Hama & Leow, 2010). Thus, it is important to investigate the relationship between learning and attention in environments that are more naturalistic than those typically studied in psychology experiments (Williams, 2013). The majority of studies have investigated the NH by examining the presence or absence of self-reported awareness (see, for example, Hama & Leow, 2010; Leow, 1997; Mackey, 2006; Williams, 2005). This is most commonly carried out by means of Think Aloud protocols (TAs). However, the use of verbal reports has been repeatedly criticised on concerns of 'veridicality' (Egi, Adams & Neuvo, 2013). This refers to the extent to which a verbal report forms a valid representation of cognitive processing. One concern is reactivity, which is the possibility that the act of reporting influences participants' cognitive processes during the task. In concurrent TAs, reactivity occurs if the act of reporting influences the way participants process and complete the task. In retrospective TAs, reactivity occurs if verbalizing the information may change the recollection. Since retrospective think-alouds introduce a delay between the task being performed and the verbalisation, there may be a tendency to reconstruct, rather than report actual cognitive processes (Bowles 2008; Leow & Morgan-Short, 2004; Sachs & Polio, 2007). Another concern which has been voiced regarding TAs is time-based decay. "Without reactivation, the newly created memory trace will fade away. Thus, a delayed, product-oriented measure of intake may be invalid as a measure of noticing, as it may be unable to detect whether noticing occurred" (Godfroid, Boers & Housen, 2013, p. 486).

"Many researchers now agree that online measures are the more reliable option" (Godfroid, Housen & Boers, 2010, p. 174) because eye tracking experiments do not face potential time decay or reactivity issues and do not only register the location of attention, but can also measure the

duration of the attention, assuming there is a relationship between where we look and what we are attending to (see below for a discussion of Just & Carpenter's (1908) 'eye-mind link'). Data that are collected via eye tracking during silent reading are not contaminated by memory demands, processes of articulation, or conscious strategies associated with the Think Aloud activity.

Rayner suggested that looking at eye movements and where visual attention is focused could be helpful in understanding comprehension problems in older learners whose natural reading behaviour is typically silent (Rayner et. al., 2006). Smith (2012) proposed that eye movement data could potentially be valuable in helping to determine "which features of the input are likely to be noticed and which are not, since we can see precisely what learners view and arguably attend to" p. 72). The premise is that eye tracking can allow us to make precise moment-by-moment inferences about the nature and amount of processing being applied without significantly altering the normal characteristics of either the task or the presentation of the stimuli (Dussias 2010). However, as Spinner, Gass and Behney (2013) pointed out, "methodology matters" and even the smallest changes in the arrangement of a text, font size and other display characteristics can lead to different results. Mangen, Walgermo and Brønnick, (2013) found that subjects who read texts on paper performed significantly better than subjects who read the on the computer screen. When reading on screen, scrolling is inevitable unless the text is within the screen size. Scrolling is known to hamper the process of reading, by imposing a spatial instability which may negatively affect the reader's mental representation of the text and, by implication, comprehension (Baccino, 2004; Eklundh, 1992; Piolat, Roussey, & Thunin, 1997). In my studies, by presenting only two sentences on a screen, I intended to minimize the potentially negative effects of scrolling.

It must be noted, however, that using eye tracking to investigate cognitive processes is based on the important assumption that there is a relationship between where we look and what

we are attending to. Just and Carpenter (1980) formulated the influential eye-mind link, "according to which there is no appreciable lag between what is fixated and what is processed" (Holmqvist et al., 2011, p. 378). If this assumption is correct, when readers look at a word, they also process it for exactly the same amount of time as the recorded fixation. Duchowski (2007, p. 205) stated that eye tracking provides "objective and quantitative evidence of a participant's visual (and overt) [his parentheses] attentional processes", but herein lies a problem. Covert attention is defined as paying attention without moving the eyes; overt attention is defined as selectively processing one location over others by moving the eyes to point at that location (Findley & Gilchrist, 2008). Therefore, a caveat in the use of eye tracking in reading studies is that although eye tracking data is precise and revealing in many ways, it is by nature limiting as we are required to make inferences about attention based on the learners' eye movements (Smith & Renaud, 2013, p. 162). Therefore, eye tracking should be treated as indicative of cognitive processing rather than a true and full reflection of it (Reichle, Warren & McConnell, 2009).

In their study using eye tracking technology, Godfroid, Boers and Housen (2013) gauged learners' noticing of new words in written input, specifically instances of focal attention. The participants in the study read twenty short paragraphs while an eye tracking machine recorded their eye movements. Eye movement analyses were conducted on participants' first fixation, fixation duration, and total time reading (see section 2.1.7.1 for explanations of eye tracking terminology). The researchers reported that they were able not only to register the locus of attention but also the duration of attention. The key finding of this study is that there is a "direct, positive relationship between the amount of attention and amount of learning" (op.cit. p. 38). Godfroid, Boers and Housen (op.cit.) claimed that this study complements and supports the NH by showing that not only more awareness, but also more attention, leads to more learning. Smith and Renaud (2013)

used eye tracking to explore the relationship between second language recasts and noticing and learning during computer-mediated communication. They report that eye tracking methodology allowed them not only to explore whether, but also how long, learners attended to the corrective targets.

The noticing hypothesis is based mainly on the claim that learners must attend to and notice linguistic features in the input, if these forms are to become intake for learning. Numerous studies have attempted to measure noticing, but the vast majority of these investigations have made use of self-reported awareness whose limitations have just been discussed. In order to address the issue of veridicality, I have used eye tracking technology to investigate the attentional constituent of noticing, as reflected in the participants' fixations and saccades. As it is reading at the word level which appears to cause great difficulty for Arabic L1 learners of English, both Study 1 and 2 will attempt to measure attention at the word level.

3.3 Conclusion

In this chapter, I described the five different types of reading interventions which I chose to include in Study Two: textual enhancement, phonological awareness training, training in word recognition and automaticity, training in oral text reading fluency and spelling instruction These interventions were incorporated into regular EFL classrooms by regular classroom teachers with students typical of the Gulf region. The design, therefore, addresses Winke's (2013) appeal that studies be conducted in conditions more representative of real classrooms and also ensures that the results are generalizable to other students and classrooms in the Gulf. To my knowledge, there have been no studies conducted which investigated the effects of a barrage of interventions on the reading proficiency of Arabic L1 EFL learners.

I have also briefly discussed the Noticing Hypothesis, and my focus on the attentional component of Noticing. The use of eye tracking technology to investigate attention removes the potential time decay or reactivity issues inherent in off-line measures and removes the responsibility from the participants to elucidate their noticing experience. Rayner, Chace, Slattery and Ashby (2006,) suggested that potential sources of reading difficulties could be identified by monitoring the eye movements of ESL/EFL learners to observe whether they exhibit the patterns observed in skilled English L1 readers. To my knowledge, Study One is the first study to compare the eye movements of Gulf Arabic L1 EFL learners with skilled English L1 speakers to ascertain how similar or dissimilar the eye movements of the two groups are. It is also the first study to compare the attention paid to consonants and vowels by English L1 and Arabic L1 EFL students.

In Chapter 4, I will describe Study One and address the research question:

Do the eye movements of Arabic L1 students at a technological college in Qatar differ from those of skilled English L1 readers while reading English sentences? If so, how do they differ and to what extent?

Chapter 5 will discuss Study Two which addresses the following research questions:

1) Do reading interventions in the EFL classroom influence the eye movements of Arabic L1 students while reading English texts? If so, how and to what extent?

2) Do reading intervention activities in the EFL classroom have an effect on students' overall reading comprehension? If so, how and to what extent?

Chapter 4. Study One: An eye tracking study of English L1 and Arabic L1 speakers reading English sentences

Introduction

In this chapter, I will describe Study One, an eye tracking experiment which compared the eye movements of skilled English L1 speakers and Arabic L1 EFL learners reading English sentences. The aim of the study was to ascertain the extent to which the Arabic L1 EFL learners' eye movements approximated those of the English L1 skilled readers. This comparison was carried out as an important preliminary phase of two studies with a broader aim to evaluate intensive reading interventions in EFL classes in a technological college in the State of Qatar. After a detailed description of the study itself, I will present the findings of the 11 metrics which were calculated.

4.1 Introduction to Study One

Reading comprehension can be viewed as involving various reading competencies. As mentioned in Section 2.1, at the word reading level, the component processes encompass letter-sound knowledge, accurate word decoding and automaticity in decoding. At the language comprehension level, the processes include activating word meanings, understanding sentences, making inferences, comprehension monitoring and understanding text structure. This "multifaceted nature of reading makes comprehension skill a sensitive barometer of overall reading development" (Rayner, Chace, Slattery & Ashby, 2006, p. 241). However, a low reading comprehension score does not give any information about which underlying difficulties contribute to it. There are very few truly diagnostic second and foreign language tests (Alderson, Brunfaut &

Harding, 2015). Therefore, teachers often lack crucial knowledge to guide them in classroom interventions intended to remediate reading comprehension difficulties. As Alderson, Brunfaut and Harding (op.cit.) pointed out, in addition to a diagnostic instrument, a diagnostician who can capably make a diagnosis is also needed. Rayner et al. (op.cit.) suggested that potential sources of reading difficulties could be identified by monitoring the eye movements of ESL/EFL learners to observe whether they exhibit the patterns observed in skilled English L1 readers. These comparisons could be useful because as the learners become more accurate in decoding words and gain in their general reading proficiency, their eye movement behaviours change substantially. For example, there is a decrease in average fixation duration, an increase in average saccade length, and a reduction in both regression rates and word skipping (McConkie et al., 1991). Contrasting the reading patterns of English L1 readers with those of EFL learners could potentially generate useful information as to how 'close' or 'far way' the learner's reading behaviour is to that of a typical skilled reader. "Although diagnosis and treatment are separate areas, diagnosis is intended to lead to treatment, and the more specific the diagnosis can be, the more likely it is that useful teaching and learning materials can be devised" (Harding, Alderson & Brunfaut, 2015, p. 326). Then, by logical extension, pedagogical interventions could be created to assist the students to approximate the eye movements of skilled readers with the goal of improving their reading at the word level.

Study One addressed the following Research Question: Do the eye movements of Arabic L1 EFL students at a technological college in Qatar differ from those of skilled English L1 readers while reading English sentences? If so, how do the eye movements differ and to what extent? Fixations, forward saccades, and regressions were investigated using data from 11 eye tracking metrics. Specifically, these metrics are 1) number of fixations 2) median fixation duration 3)

forward saccade length 4) number of forward saccades per individual 5) length of regressions and 6) proportion of saccades that are regressions. The first 6 metrics were chosen as they have been shown in the literature to vary with reading proficiency and experience (see Brunfaut, 2016; Brunfaut & McCray, 2015; Holmqvist et al., 2011). The last 5 metrics were chosen because they are of particular relevance to this study and to an examination of the 'vowel blindness' phenomenon discussed in Section 2.6. The particular metrics that were calculated were: 7) sum of all visits on consonants, 8) mean of all visits on consonants 9) sum of all visits on vowels 10) mean visits on vowels and 11) mean proportion of visits on vowels.

4.2 Study one

Methodology

4.2.1.1 *Participants*

There were a total of 75 participants in the study consisting of two groups. The English L1 participant group comprised 36 faculty and staff members of a technical college in the State of Qatar. 17 of the participants were males and 19 were females. Participation was voluntary and solicited through personal contact. All were Canadian, English L1 speakers, between the ages of 27 and 64 with normal to corrected-to-normal vision. There were 6 participants in the 25-35 age range, 12 in the 36-46 range and 18 in the over 46 range. All had obtained at least one university degree, and although university degree is arguably a crude measure of reading proficiency, the participants could be considered as more skilled than the general population of readers. No formal information was elicited regarding learning differences or dyslexia. Each participant read and signed a consent form, approved by the technical college and by Lancaster University. The consent form reiterated the voluntary nature of the study and explained that the researcher was interested

in reading behaviours in general (see Appendix A for a copy of the consent form). All participants were informed of the exact nature of the study immediately upon completion of the experiment and were shown their own eye movement recordings if they were interested. All personal data and results were anonymised.

The Arabic L1 EFL group was a convenience sample of 39 students from the same technical college in Qatar. They were students from five sections of FL 1080, an intensive EFL course designed for students at a B1 level of the Common European Framework of Reference (CEFR). There were 28 female and 11 male participants ranging in age from 18 to 26. There were 20 participants in the 18-20-year range, 15 in the 21-24 range and 4 in the over 24 range. All were Qatari nationals who had progressed through the same educational system in Qatar. All participants had normal or corrected-to-normal vision. No information was elicited regarding dyslexia or specific learning difficulties. At the time of this study, there were no formal procedures, either in the school system or at the college where the study took place, to diagnose dyslexia. Cultural issues prevented students from reliably self-assessing any learning differences.

Background information and results for all participants were anonymised. Each participant signed an informed consent form which was explained in detail with the help of one male and one female Arabic/English bilingual student assistant. In exchange for participation in the study, the participating teachers deducted one low quiz/ assignment from the students' final grade. Permission for this incentive was kindly granted by the Dean of Language Studies and Academics at the technical college and the five teachers of the classes involved agreed to this arrangement. This particular incentive provided great motivation resulting in 100% participation from the students in the four classes.

4.2.1.2 *Apparatus*

Reading materials were presented on a 23-inch monitor with a screen resolution of 1920 x 1080 pixels. The monitor was attached to an HP Z400 Workstation PC interfaced with a Tobii TX 300 eye tracking system which is equipped with a large head movement box. A head movement box is "the volume relative to the eye tracker in which a participant can move without compromising the quality of the recorded data" (Holmqvist et al., 2011, p. 58). The large head box allowed the participants to move during the eye tracking sessions while maintaining accuracy and precision at a sampling rate of 300 Hz. This allowed eye movements to be recorded without using a chinrest, thus boosting the ecological validity of the study. Window blinds, in conjunction with overhead and free-standing lighting, ensured indirect lighting at a consistent Lux of 500. This precautionary measure was taken because direct sunlight contains enough infrared light to outshine the infrared illumination, resulting in complete and immediate data loss (Holmqvist. et al., 2011). The standard of room illumination was kept constant for all participants to ensure that time of day or weather conditions did not influence the lighting. Each participant sat between 63 and 65 cm from the monitor in a comfortable stationary chair.

All recordings were carried out in the researcher's office which is located in a relatively quiet section of the technological college. The small vertical window in the office door was covered with dark paper and all signs, pictures and notices were removed from the wall behind the eye tracker. A large sign on the door advised any potential visitors not to knock or enter the room, as recording was in progress. In this way, visual and auditory distractions were kept to a minimum.

4.2.1.3 *Materials*

A series of 25 signs was initially presented to two experienced EFL teachers who were not involved in the study. They carried out visual stimuli ratings and chose 16 of the signs which they considered to require the lowest level of visual literacy. Participants were subsequently presented with these 16 images which were common everyday signs such as street signs, shop signs, school signs etc. The signs themselves were not marked for analysis in Tobii Studio as their sole purpose was to provide a plausible reason for the participants to pay attention and read the two sentences which followed each sign. Participants were instructed to look at each sign for as long as they wished, and when they were ready, they were to press the space bar, whereupon a new screen would appear with two sentences written on it (see Appendix B for examples of the signs and their 2 corresponding sentences).

Participants were asked to silently read the two sentences at their own pace and tell the researcher which sentence referred to the sign they had just read. Silent reading was chosen because processing is uncontaminated by memory demands or articulation processes (Rayner et al. 2006). Also, eye movements differ somewhat for reading silently versus reading aloud. When reading aloud, mean fixation durations are longer than in silent reading, and the eyes tend to get ahead of the voice; consequently, there are many fixations in which the eyes appear to be holding in place so as to not get too far ahead of the voice (Rayner, 2009). For this reason, the vast majority of eye tracking research on reading is for silent reading (Rayner, 1998). Participants could not go back to the previous screen. As all signs were visually straightforward and those with text had fewer than 7 words, there was little chance that participants would forget the content of each sign.

Sentences were typed in black in a 36 font on a pale yellow background in Courier New. Courier New was chosen because it is a mono-fixed font which facilitated the uniform creation of the Areas of Interest (AOIs) which are user-defined areas in a stimulus for measuring viewing behaviour. It was not possible to counterbalance the stimuli, as each sign was connected to the two sentences which followed, and it was not possible to separate them in the Tobii studio software unless two separate experiments were designed.

Each pair of sentences was matched as closely as possible for length. Additionally, two experienced EFL teachers, not involved in the study, were asked to rate the pairs for comparable complexity. The teachers' feedback was incorporated into the final draft of the sentences. Word length was held as constant as possible. This is important because it affects word skipping. Short words might be skipped because they are processed parafoveally and therefore, they may not elicit sufficient data to conduct meaningful statistical analysis (Jegerski & VanPatten, 2013). However, short function words were needed as part of the stimuli, precisely because skipping behaviours were of interest. Therefore, considerable effort was expended on choosing signs which would allow the researcher flexibility in selecting the length of words in the pairs of sentences. Again, using feedback from the same two experienced EFL teachers, I endeavoured to create sentences which did not contain words that might be unknown to the Arabic L1 EFL participants, otherwise the task might have proven demotivating. As the task consisted of paired sentences and not a coherent, cohesive text, a CohMetrix L2 (Graesser, McNamara, & Kulikowich, 2011) readability index was not computed. Instead, the traditional approach for scaling texts with a single metric of text ease or readability was calculated. These measures are robust predictors of sentence-level understanding (McNamara, personal communication, 2014). Flesch-Kincaid Grade Level scores

were calculated and these provided confirmatory evidence that the pairs of sentences were of comparable readability levels.

Table 4.1 below illustrates the features of the 16 pairs of sentences.

Table 4.1 *Features of the 16 pairs of sentences*

Feature	**Sentence 1**	**Sentence 2**
Minimum FK Grade level	2.2	2.3
Maximum FK Grade Level	6.2	6.2
Mean FK Grade Level	3.831	3.993
SD	1.25	1.21

Table 4.2 shows the FK Grade level for each of the sentences

Table 4.2 *FK Grade level for each of the sentences*

Sentence #	**Sentence 1**	**Sentence 2**
1	2.3	2.3
2	4.4	4.4
3	4.4	3.9
4	5.2	5.2
5	3.6	3.7
6	3.6	3.7
7	2.2	2.3
8	4.4	4.4
9	2.3	2.4
10	2.4	4.4
11	3.7	3.6
12	5.2	5.6
13	4.8	4.8
14	4.4	4.7
15	2.2	2.3
16	6.2	6.2

Table 4.3 below shows the two groups'performance on the sign comprehension test

Table 4.3 *The two groups'performance on the sign comprehension task*.

	English L1	**Arabic L1**
High	16	15
Low	12	6
Mean	15.14	11.11
Median	16	11.5
SD	1.06	2.21

4.2.1.4 *Procedure*

The researcher and two bilingual English/Arabic student-helpers visited the five EFL classes to show a short video on the Tobii TX 300 and explain, in general terms, the nature of the study. The presentation was carried out in both English and Arabic. Participants were told only that the purpose of the study was to investigate how Arabic L1 EFL students read in English. At that time, consent forms were explained and then signed by all participants. Then, before the actual recordings, several pre-trials with non-participant EFL students were conducted to ensure that explanations and instructions were clear. The instructions were consequently modified to eliminate any confusing information. Instructions for the experiment were in English and were the same for all participants. Instructions were written in the form of a script and read aloud by the researcher to ensure that each participant was given exactly the same directions and information. This safeguard was implemented as research shows that "task instruction has a very strong influence on eye-movement behaviour" (Holmqvist et al. 2011, p. 134). If participants have different

conceptions of the task, they are likely to behave differently. The pre-trials also ensured the smooth running of the equipment in addition to confirmation that instructions were clear. Explanations and instructions were presented individually in the researcher's office for the English L1 participants.

A 9-point calibration session was carried out before each participant was recorded. Calibration refers to the procedure of data collection so that the "coordinates of the pupil and one or more corneal reflections in the coordinate system of the eye video can be converted to x and y coordinates that represent the participants' point of regard in the stimulus space" (Blignaut & Wium, 2013, p. 67). Calibration speed of the 9 points was set at medium. The calibration dots were displayed in red on a grey background, and calibration was set to use the full screen. In a calibration session, participants look at the nine different points and the eye tracking software samples a few hundred milliseconds of data for each individual. Calibration sessions are extremely important as there are several possible difficulties which may interfere with the integrity of the data. Participants are often miscalibrated simply because they happen to look in the wrong direction or they fixate for only a short time on several of the points. As recommended by Holmqvist et al., (2011), all participants were instructed to keep their gaze fixed in the centre of each point until it disappeared. There are also other factors which may interfere with the validity of the calibration and subsequent data. For example, the wearing of mascara poses a problem for data integrity. The software that identifies the pupil is confused by the other dark area in the vicinity of the pupil and locks onto the mascara rather than to the pupil (Holmqvist et al., op.cit). Long, thick eye lashes can also be problematic as the real corneal reflection is partially obstructed underneath the lashes. In anticipation of students with long eye lashes or copious amounts of mascara, an eye lash curler and mascara remover were on hand in the researcher's office. Eye glass cleaner was also available

for those participants whose smudgy glasses might have interfered with the calibration. Calibration was considered valid if data could be captured for each dot. When necessary, calibration was repeated for one or more dots. A second calibration was carried out on two participants because of the position of their eyeglasses. After raising the eyeglasses higher on the bridge of the nose, the results of the calibration were much improved. One participant was recalibrated successfully wearing his regular reading glasses instead of his progressive eye glasses. Holmqvist et al., (2011) noted that bi-focal eye wear may introduce a border "in the midst of the eye video that makes calibration difficult and recording close to impossible" (p. 124). One participant was recalibrated the following day wearing eye glasses instead of his soft contact lenses. Holmqvist et al. (op. cit.) explain that as a result of an imperfect fit between the eyeball and the lens, tiny air bubbles collect underneath the soft contact lens. When the infrared illumination is reflected in the bubbles, the light is fragmented into a number of reflections and the eye tracker will randomly choose any of the reflections as the corneal reflection. This results in inaccurate data samples known as "optic artefacts" (Holmqvist op.cit. p. 124). The subsequent recalibrations produced high quality data highlighting the importance of calibration conditions. Only one participant was excluded after calibration and this was because of 'droopy eye lids' (Holmqvist, et al. op.cit). Droopy eyelids cover the pupil in the lower gaze directions and at a certain gaze angle, the pupil may be entirely covered which will result in complete data loss (Holmqvist, op. cit.).

For the Arabic L1 participants, recordings were carried out in the first week of the semester before any reading instruction had taken place. Recording sessions, including instructions, took approximately 40 minutes for the Arabic L1 EFL participants and 20 minutes for the English L1 participants. Both eyes of each participant were tracked and analysed.

4.2.1.5 *Data Analysis*

After all the recording sessions were completed, the quality of the data was investigated. This was estimated by visual inspection of the gaze plots for the 16 sets of sentences for all participants. Three participants (4% of the total number of participants) were eliminated at this stage because of poor quality data. This ensured that the remaining recordings were of high integrity. *Figure 4.1* below shows an example of good quality data in which the fixations are more or less right on the word and on the lines of print. *Figure 4.2* shows an example of poor quality data in which the fixations do not appear to be on the lines of print. Holmqvist et al. (2011) report that in their experience around 2-5% of the data from an average population of non-screened European participants needed to be dismissed due to participant–specific tracking difficulties, but they note that this number varies significantly, with some studies reporting losses of 20-60% of participants/trials.

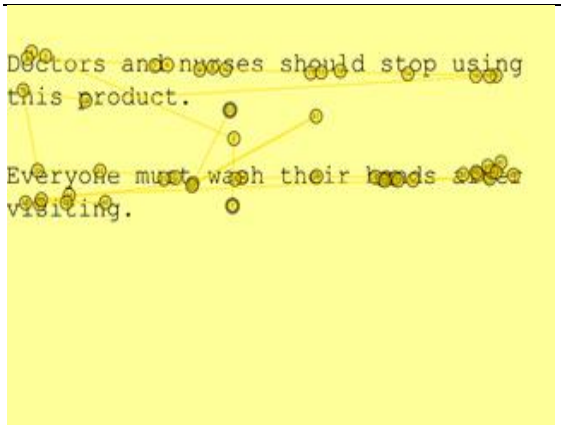


Figure 4.1 Good quality data

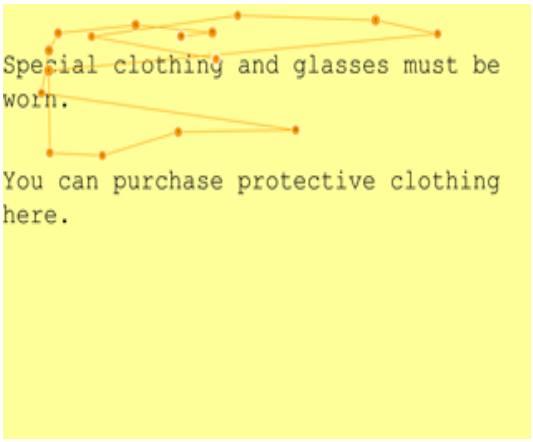


Figure 4.2 Poor quality data

Data from the first sign the participants looked at, and its two related sentences, were eliminated from data analysis. This constituted a warm-up which allowed participants to become familiar with the Tobii eye tracker, the format of the stimuli, and the demands of the task.

One important aim of analysing eye tracking data is to find patterns and gain new knowledge from the data. However, discovering new knowledge necessitates the capacity to discover any unexpected patterns (Yu, Yurosky, & Xu, 2012). As all data analyses could not be specified a priori, visual techniques were implemented to allow for detection of any interesting or unexpected trends. Visual inspection of eye movement data can provide "critical insights to guide us to determine which aspects of the data to focus on" (Yu et al., op.cit, p. 37). Therefore, aggregate heat maps were generated in Tobii studio for all recordings for the English L1 and Arabic L1 groups. Heat maps are two-dimensional graphical representations of data in which the values of a variable are shown in colours. Red is warmer than orange, orange is warmer than yellow, yellow is warmer than green etc. The amount of heat that is shown is proportional to the level of the represented variable. The assumption is that the amount of heat represents the amount of attention given to a particular area or word on the screen (Tobii White Paper, 2010). From the three different

types of heat maps available (count, absolute duration and relative duration), relative duration was chosen as this 'evens out' individual performance. That is to say, each participant's data is given the same weight. For example, if one participant spends much more time on the task than the other participants, the relative duration heat map does not weight his/her data more heavily. Heat maps were generated for the 15 sets of sentences for both groups. As was previously mentioned, the first set of sentences was used as a warming-up task and data from it were not included in the analyses. *Figure 4.3* shows a typical example of the English L1 group fixation durations. *Figure 4.4* shows a typical example of the Arabic L1 group's median fixation durations.

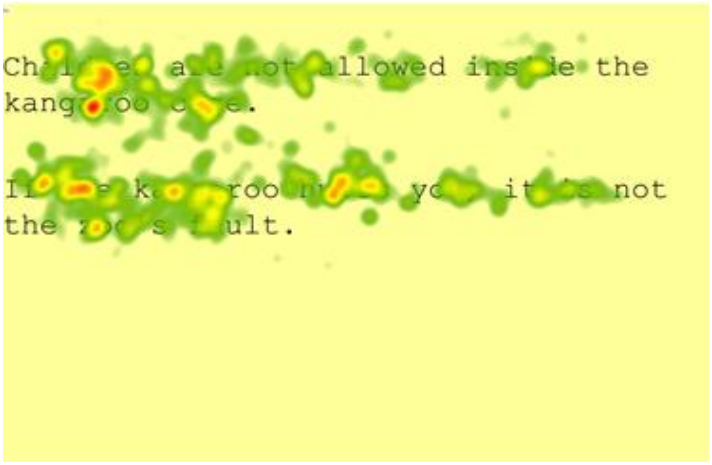


Figure 4.3 English L1 median fixation durations

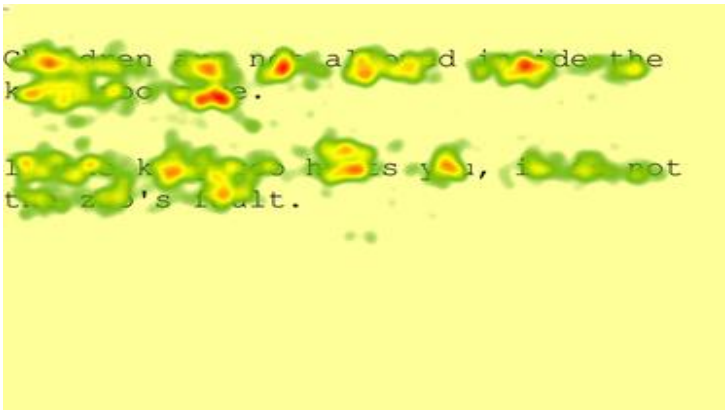


Figure 4.4 Arabic L1 median fixation durations

The general viewing behaviour seen in the heat maps shows that the English L1 speakers appeared to fixate more on content words, less on function words and to skip some words altogether. This is consistent with previous research on proficient English L1 readers (Rayner, 1998). The pattern contrasts sharply with the viewing behaviour of the Arabic L1 readers who seemed to fixate on each individual word. There is no evidence of content words being fixated more than function words and it appears as if there is no word skipping at all. The heat maps give the impression that even the shortest function words were fixated.

Gaze plots were then generated for both groups on all recordings. A gaze plot is a representation of a reader's eye movements across the screen, fixation by fixation, by means of dots. Each dot of the Gaze Plot is drawn when a fixation is registered at a certain point on the screen. Each fixation is represented by a dot. The dots are numbered according to the sequence of the fixations (Tobii technology, 2014). *Figure* 4.5 shows a typical example of a gaze plot for the English L1 group. *Figure 4.6* shows a typical example of a gaze plot for the Arabic L1 group.

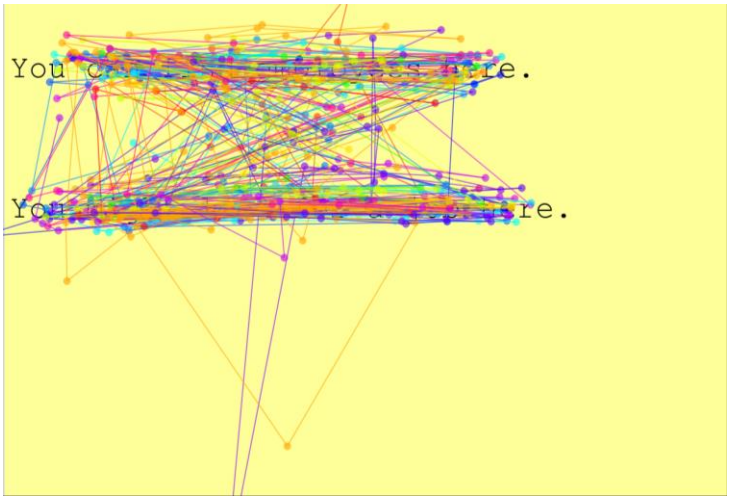


*Figure 4.*5 English L1 fixations

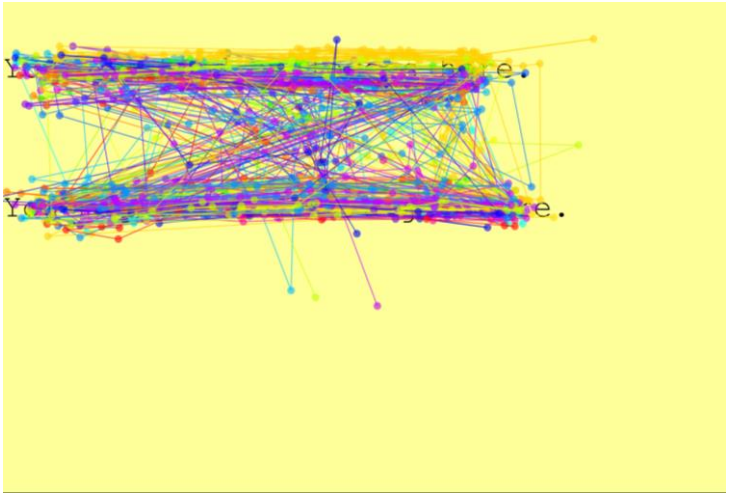


*Figure 4.*6 Arabic L1 fixations

The general viewing behaviour in the gaze plots provided unexpected patterns. It appeared that the Arabic L1 speakers switched between sentences appreciably more than the English L1 speakers in order to confirm comprehension. It also appeared as if that they made more regressions and forward saccades within the sentences. However, these visual observations needed a more in-

depth analysis. Visualizations such as heat maps and gaze plots provide a straightforward and intuitive synopsis of large data sets but this is also their disadvantage. Because of their simplicity, it is tempting to draw unsupported and invalid conclusions based on them. Such visualizations do not provide any method for systematic and statistical comparisons between conditions (Holmqvist et al., 2011). They are helpful to characterize the general viewing behaviour rather than to explain it. They can point out the regions that attracted participants' gazes but it is "highly speculative to draw any conclusions of what made people look there" (Holmqvist et al. op.cit., p. 241).

The valuable information provided by the gaze plots led to a re-examination of the metrics intended for statistical analyses and it was decided that in addition to fixation measures, metrics pertaining to forward saccades and regressions would be included. The software used, Tobii Studio version 2. X did not calculate forward saccades and regressions; therefore, the raw data for the AOIs around the sentences were exported into R (2012) to ascertain if there were any significant differences between the two groups of participants in terms of both their fixation and saccadic patterns. Fixation data from the AOIs around the vowels and consonants were calculated in Tobii Studio 2. X according to the original plan. The shapes of the AOIs are defined standard shapes available in the Tobii Studio toolbox: ellipses, rectangles and polygons. The AOIs for this study were manually drawn around each of the 32 sentences (see *Figure 4.7*). Another set of AOIs was drawn around each vowel and each consonant in every sentence so a separate analysis could be conducted (see *Figure 4.8*).

According to Holmqvist et al. (2011) the AOIs are part of the hypothesis one is investigating because they determine which areas in space the fixation and transition data should be calculated on. This has two important implications. If one modifies the AOIs, the research hypothesis also changes. If one creates the AOIs after data recording, one is making post-hoc

hypotheses. Study One addressed the following Research Question: Do the eye movements of Arabic L1 EFL students differ from those of skilled readers while reading English sentences. If so, how do the eye movements differ and to what extent? In asking this question I was not only interested in the general viewing behaviours of the two groups but also in the differences in the allotment of visual attention when comparing the eye movement on vowels. In order to do this, two separate sets of AOIs needed to be created: one at the sentence level and one at the individual letter. Otherwise, no claim could be made about the participants' viewing behaviour on vowels and consonants specifically.

With reading stimuli, AOIs can be drawn around sentences, words and morphemes. However, a question that is often asked is, how big these AOIs should be? "Smaller AOIs increase selectivity and ignore more extraneous gaze at the risk of losing some valid looking. On the other hand, larger AOIs increase sensitivity and may include more extraneous data potentially capturing more of the actual looking at the target" (Tobiipro, 2017, p. 2). It is up to the researcher to choose a suitable balance between selectivity and sensitivity. "Taking both accuracy and precision into account, the practical minimal size of an AOI is around $1\text{-}1.5^0$ for high-end eye trackers because this is the size of the fovea and the best eye trackers have the precision to accommodate such a size" (Holmqvist et al., 2011, p. 223). This corresponds to 1.2 to 1.8 centimetres on a computer screen viewed from a 66.6 centimetre distance.

The assignment of AOIs to the consonants and vowels was straightforward. However, visual intake and recognition often complicate the AOI division. Holmqvist et al. (2011) note that if the stimulus is simple and the AOIs are so close together that the readers can take in one AOI in peripheral vision while looking at the other, it is questionable to contrast dwell times from the two areas and claim that the visual intake from one AOI is larger than from the other. However, a

phenomenon known as ‘crowding’ explains that “as peripheral information becomes more cluttered, it is difficult to distinguish between different elements away from the current point of fixation. Therefore, for complex displays, AOIs which are close together may not cause a problem because crowding restricts focus to the fovea (Holmqvist et al. op.cit, p. 217).

Another issue which needed to be addressed was the space or margin between the AOIs. Adjacent AOIs should have sufficient spacing between them to allow the desired balance of specificity and selectivity by applying the “1 degree” guideline. When designing stimuli and AOIs, the “1 degree" guideline is an easy way to take into account the extent of the foveal field (Tobiipro, op.cit). Considerable attention was therefore paid to the positioning of the AOIs. Exact positioning is crucial because it can determine whether a significant effect is revealed or not. AOIs “should not overlap because of the danger that single AOI hits and transitions will be counted twice, rendering the statistics difficult if not impossible to interpret” (Holmqvist et al., 2011, p. 221).

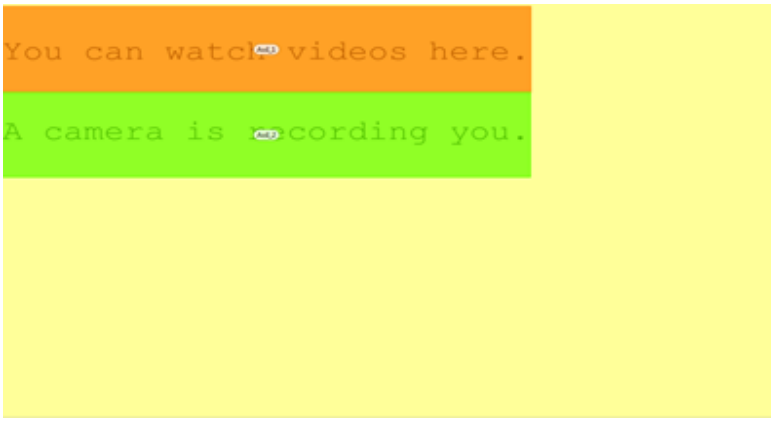


Figure 4.7 AOIs around the sentences

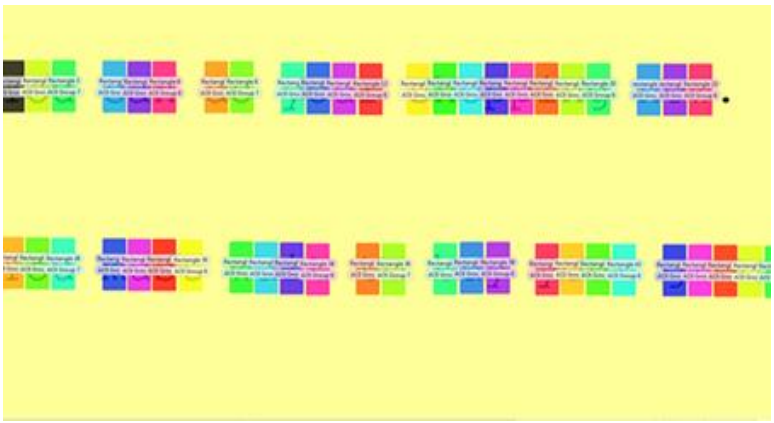


Figure 4.8 AOIs around vowels and consonants

4.2.1.6 *Metrics*

In every eye tracking experiment a decision must be made regarding which metrics to employ. "Finding the right measure is often the most difficult thing when designing an experiment, especially for someone who is not trained in mathematics, computer science, or experimental psychology" (Holmqvist et al. 2011, p. 455). There are over 120 measures for eye-movement data listed in Holmqvist et al. (op.cit) and the decision of which ones to use depends upon the hypothesis and research design. As an example, in many eye movement experiments investigating reading, researchers build experimental sentences which have been created to cause participants difficulty while they are reading, and control sentences which are usually similar but are designed so they do not cause the readers problems. The researcher then compares reading times for regions of the experimental sentences with reading times for the control sentences and in this way is able to detect the degree of disruption to normal reading that the experimental manipulation caused. When measuring processing difficulty in this way, researchers generally report three reading time measures: first fixation duration, first pass reading time and total reading time (Liversedge,

Paterson and Pickering, 1998). Additionally when researchers are interested in identifying critical regions of text, they examine how long it takes readers to read the regions of interest. The standard measures used here are first pass reading time, go-past duration and second pass reading time. First fixation durations are often reported when the researcher is interested in spill-over effects from the previous region. This measure is generally taken to be the very earliest point at which we might expect to see an effect due to the experimental manipulation, as this is the first time the reader has directly fixated the region in which disruption to processing is anticipated (Clifton, Staub & Rayner, 2007). As my research question addressed the differences in the eye movements in different reading populations and not the kinds of disruption which cause reading difficulty, measures such as these were not selected.

At the word and sentence level the following metrics were investigated:

1) Number of fixations

This metric is a "very good general measure" (Holmqvist et al. 2011, p. 413) and has been used as an indication of search efficiency/difficulty, experience/proficiency and also of dysfunctions/difficulties in reading. The number of fixations overall is presumed to be negatively correlated with search efficiency. A high number of fixations would indicate difficulty in interpreting the information. A low number of fixations could mean that the participant is experienced or that the task is too easy. Proficient readers make fewer fixations than beginners (Rayner, 1998) and dyslexic readers (Holmqvist et al., 2011; Prado, Dubois &Valdois, 2007).

2) Fixation duration (expressed in milliseconds.)

Eye fixation durations are widely used as measures or indicators of processing times for fixated words or other text segments during reading (Rayner & Pollatsek, 1989; Rayner, 1998). This measure "is likely to be the most used measure in eye tracking research" (Holmqvist et al., 2011, p. 377). Most researchers use fixation durations because "these are assumed to reflect perceptual intake and processing" (Holmqvist et al., op.cit. p. 379). Research has shown that how long the eyes rest on a word reflects the ease or difficulty with which that word is processed. Less frequent words, more complicated texts and complicated grammatical structures receive longer fixation durations. Longer durations are often associated with "deeper and more effortful cognitive processing" (Holmqvist et al., op.cit. p. 381).

3) Forward saccade length (expressed in pixels).

Also referred to as saccade amplitude, this metric is also widely used in reading research. In reading, saccade length is restricted by the visual span and thus must be on an average of 7-8 letters. Shorter saccadic amplitudes are an indication of increased cognitive load. Shorter amplitudes are evidenced in the eye movements of young children when they are first beginning to read. Poor readers and readers with dyslexia also exhibit shorter saccade length (Rayner & Pollatsek, 1989).

4) Number of forward saccades per individual

There is a relationship between the length of a saccade and the number of saccades one needs to read a sentence or text. That is to say, the longer the saccade, the fewer will be needed.

5) Regression Length (expressed in pixels).

Short regressions may be due to problems the reader has in processing the currently fixated word. Longer regressions occur "because the reader did not understand the text" (Rayner, 1998, p.

375). Regressions inside words are thought to reflect lexical activation processes while regressions between words reflect sentence integration processes. More proficient readers make longer regressive saccades which allow them to resolve anaphoric expression, while poor readers must search for the antecedent (Holmqvist et al., 2011).

6) Proportion of saccades that are regressions.

Because less proficient readers make shorter regressions, it is necessary for them to make a greater number of them in order to clarify comprehension.

To further investigate the presence of the 'vowel blindness' phenomenon as described in Chapter 2.6 the following metrics were investigated:

7) Sum of all visits on consonants

8) Mean of visits on consonants

According to the 'vowel blindness' hypothesis, Arabic L1 readers tend to focus their attention more on consonants than on vowels.

9) Sum of all visits on vowels

10) Mean visits on vowels

11) Proportion of vowel visits

Again, according to the 'vowel blindness' hypothesis, there is an assumption that Arabic L1 speakers "lack an awareness of the function vowels perform in English" (Ryan, 1997, p. 189) and consequently do not recognize or attend to them.

4.2.1.7 *Fixation Filters*

Fixation filters are algorithms for detecting fixations. Fixation identification is an integral part of eye movement data analysis because a good identification algorithm ensures valid fixation and saccade locations. However, different types of filters, and also the same filter with different parameters, can yield different results. Fixation and saccade output is very sensitive to the choice of the algorithm and setting. Although otherwise similar, "studies using different settings of the peak saccade velocity are not directly comparable" (Holmqvist et al. 2011, p. 158). Both Holmqvist et al. (op.cit.) and Tobii Training (2013) suggest filtering the raw data through several of the different filters available to determine the quality of the data obtained. The Tobi I-VT, which is a velocity-based classification algorithm was selected after it provided the least amount of 'noise' when viewing preliminary raw data on all participants. *Figure 4.9* and *Figure 4.10* below show examples of the same data filtered with the Tobii Studio Clear View filter and with the Tobii I-VT using its default settings.

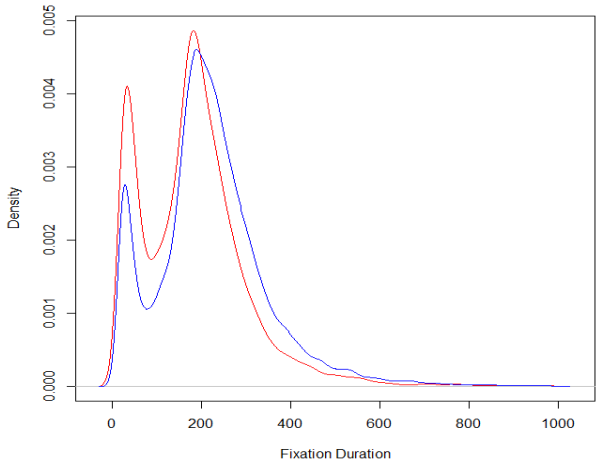


Figure 4.9 Clear View fixation filter

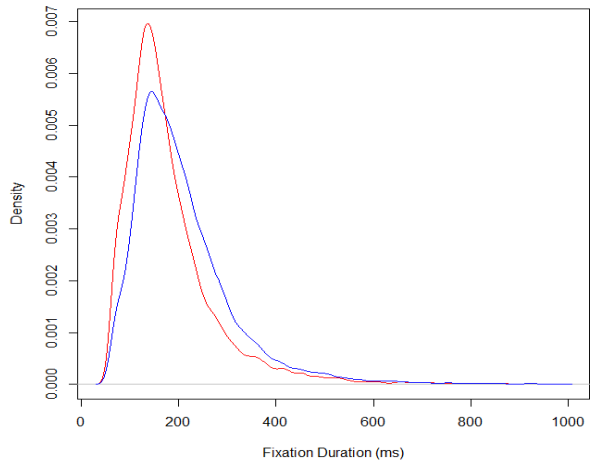


Figure 4.10 I-VT fixation filter

As can be seen, the quality of the data is quite different depending upon the algorithm. In *Figure 4.9* the peak on the left represents what is considered to be random 'noise' due to the unsuitability of the fixation filter for reading research.

Table 4.4 below shows the hypotheses and expectations of the eye tracking metrics in Study One.

Table 4.4 *Hypotheses and expectations of Study One*

Metrics calculated	**Hypotheses and Expectations**
Number of fixations	The Arabic L1 EFL participants will make more fixations than the English L1 participants.
Median fixation duration	The Arabic L1 EFL participants will make longer fixations than the English L1 participants.
Forward saccade length	The Arabic L1 EFL participants will have shorter forward saccades than the English L1 participants.
Number of forward saccades per individual	The Arabic L1 participants will execute a greater number of forward saccades.
Regression length	The Arabic L1 EFL participants will make shorter regressions than the English L1 participants.

Proportion of saccades that are regressions	The Arabic L1 EFL participants will execute a higher proportion of regressions than the English L1 participants.
Sum of all visits on consonants Mean of all visits on consonants	The Arabic L1 EFL participants will make more visits on consonants than the English L1 participants and the mean number of visits will be higher for the Arabic L1 EFL group
Sum of all visits on vowels Mean of all visits on vowels	The Arabic L1 EFL participants will execute fewer visits on vowels and the mean number of visits on vowels will be lower than for the English L1 participant group.
Proportion of vowel visits	The Arabic L1 EFL participants will execute a lower proportion of visits on vowels than the English L1 group.

4.2.2 Results and Discussion

4.2.2.1 *Results of metrics calculated on words and sentences*

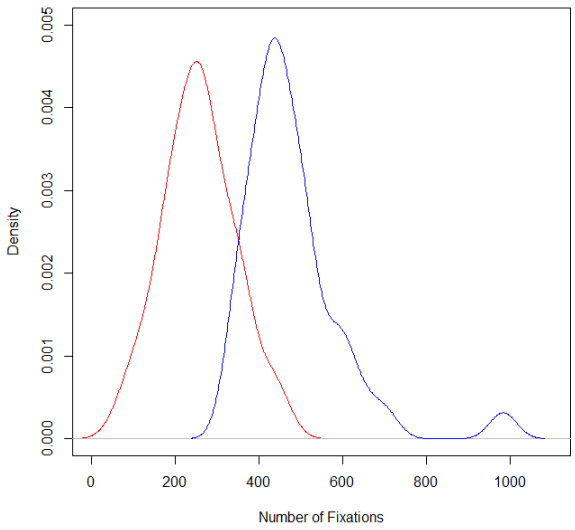
Table 4.5 below presents a summary of the outcomes of the 6 metrics calculated in R for the 33 English L1 and 39 Arabic L1 EFL participants. The data had skewed distributions; therefore, the median was used as a more appropriate measure of central tendency, as the mean is more sensitive to skew and is pulled in the direction of the extreme scores (Sirkin 2006). A Wilcoxon rank-sum test, the non-parametric alternative to the two sample t-test, was conducted, as the assumptions of roughly normal distributions and homogeneity of variance did not hold for the data. The critical value for hypothesis tests, p, was set at 0.05. The results for each metric will be discussed separately.

Table 4.5 *Summary of metrics 1 to 6 for Arabic L1 EFL and English L1 groups*

Metrics	**Median Arabic L1**	**Median English L1**	**W**	**p**
Number of fixations per individual	450	252	48	<0.001
Median fixation duration	187 ms	157 ms.	91836740	<0.001
Forward saccade length	198 px.	317 px.	47344944	<0.001
Number of forward saccades per individual	290	155	59	<0.001
Length of regressions	-156 px.	-252 px	8061788	<0.001
Proportion of saccades that are regressions	0.034	0.037	805	≤0.069

Number of fixations per individual

The heat maps which were generated earlier in the analyses gave an indication of some basic differences in the fixation patterns of the two groups. For example, it appeared that the Arabic L1 readers made more fixations than the English L1 readers. Statistical analysis in R confirmed this initial hypothesis. There is a highly significant difference (*W = 48; p =* <0.001) between the individuals in the Arabic L1 group (*M = 450*) and the individuals in the English L1 group (*M =252*) in terms of median number of fixations (see *Figure 4.11*).

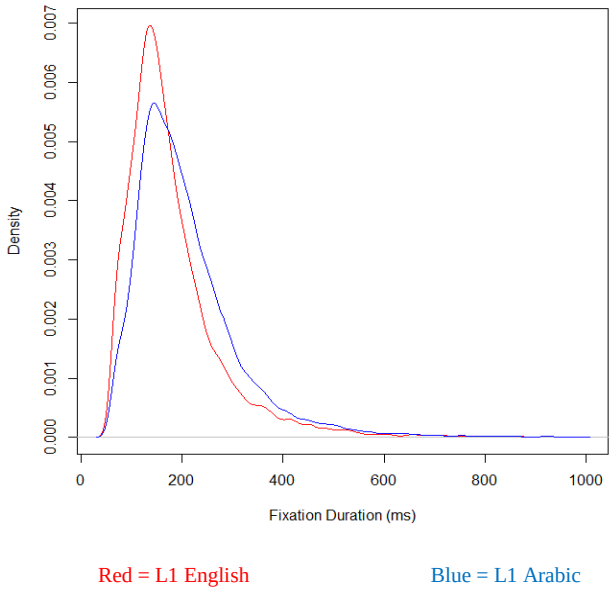


	N	Median Fixations
L1 English	33	252
L1 Arabic	39	450
Total	72	

Figure 4.11 Number of fixations per individual

Median Fixation Duration

Figure 4.12 shows the median fixation duration of the two groups. There is a highly significant difference (*W= 91836740*; *p= <0.001*) between the two groups in terms of the median length of fixation duration. In the 10,714 fixations made by the English L1 group, the median length is 157 ms. The Arabic L1 EFL group made 21,585 fixations with a median of 187 ms. This means that the Arabic L1 EFL participants not only made more fixations as a group but their fixations were longer.

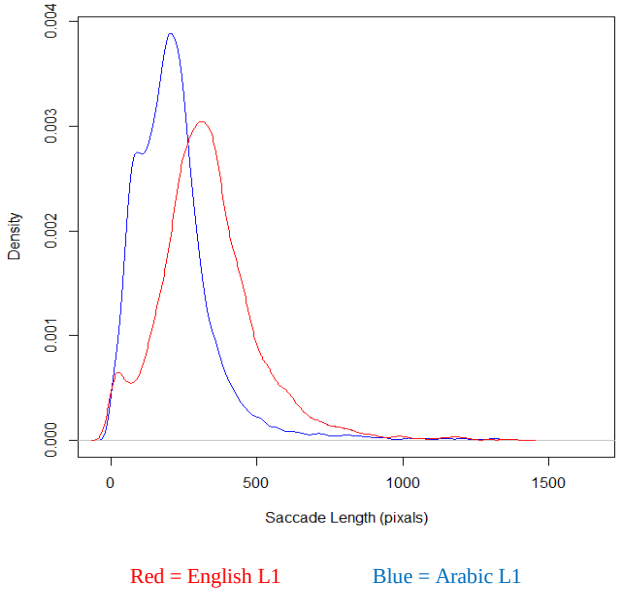


	N	Median Length (ms.)
L1 English	10714	157
L1 Arabic	21585	187
Total	32299	

Figure 4.12 Median fixation duration

Forward Saccade length

Figure 4.13 illustrates the median saccade length of the two groups. As can be seen, the median saccade length of the L1 English group (*M=317 px.*) is significantly longer (*W=47344944; <p=0.001*) than the L1 Arabic group (*M=198 px.*).

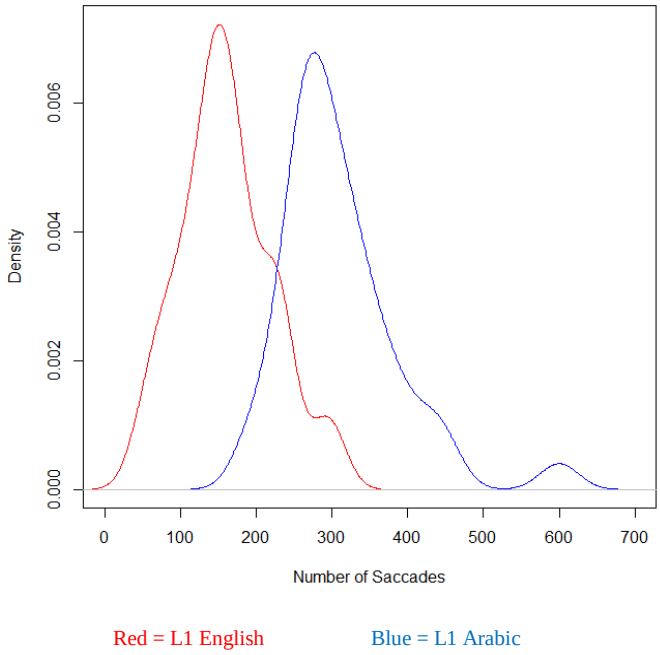


	N	Median Length (px.)
L1 English	5310	317
L1 Arabic	12081	198
Total	17391	

Figure 4.13 Forward Saccade Length

Number of Forward Saccades Per Individual

Figure 4.14 shows the median number of forward saccades per individual in each group. The median number for the L1 English individuals (*M=155*) is significantly lower *(W=59; p= <0.001)* than the median (*M =290*) for the L1 Arabic EFL participants.

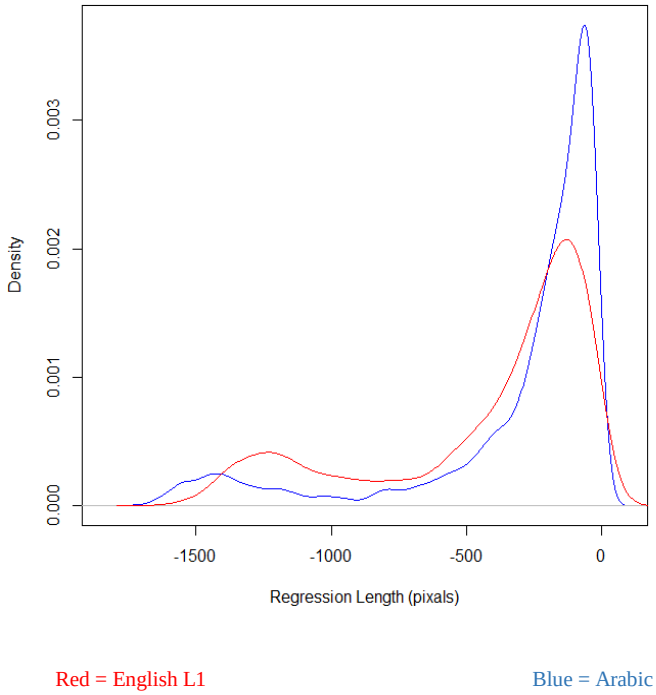


	N	Median
L1 English	33	155
L1 Arabic	39	290
Total	72	

Figure 4.14. Number of forward saccades per individual

Median Regression Length

As can be seen in *Figure 4.15*, there is a highly significant difference (*W=8061788; p= <0.001*) between the median regression length for the Arabic L1 group *(-156 px*), and the mean regression length for the English L1 group (*-253 px*.).

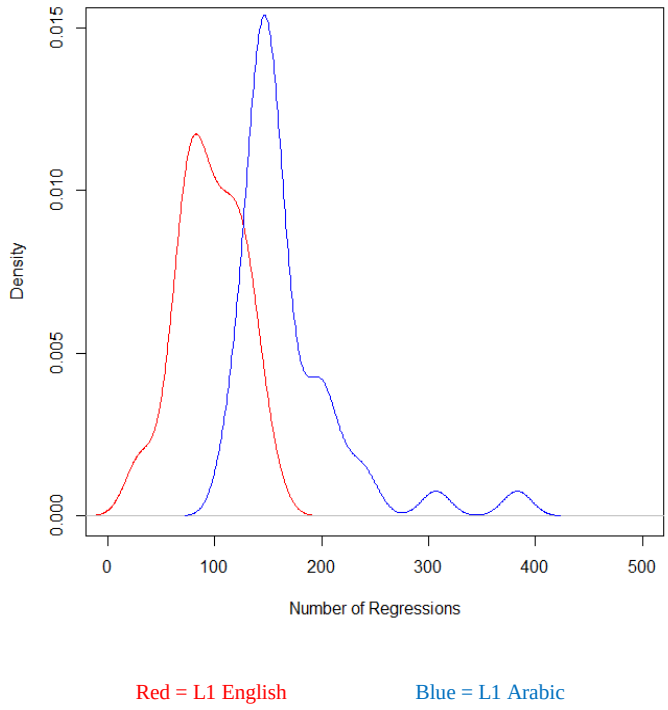


	N	Median Length (-px)
L1 English	3171	-253
L1 Arabic	6583	-156
Total	9754	

Figure 4.15 Median Regression Length

Proportion of saccades that are regressions

As can be seen in *Figure 4.16*, the difference in the proportion of regressive movements for the Arabic L1 group (*M=0.34*) and the proportion of regressions of the English L1 group (*M=0.37*) is not significantly different (*W=805; p= <0.069*).



	N	Median Number
L1 English	33	0.37
L1 Arabic	39	0.34
Total	72	

Figure 4.16 Proportion of saccades that are regressions

4.2.2.2 *Results of metrics calculated on vowels and consonants*

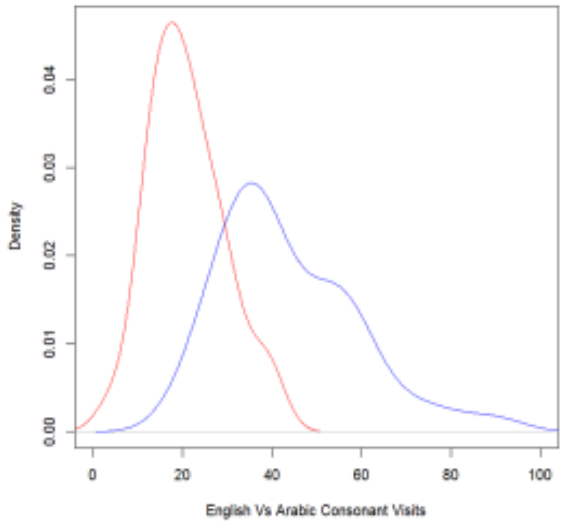
To investigate the phenomenon of 'vowel blindness' described in Chapter 2.6, a separate analysis was carried out using the data from AOIs drawn on the 483 vowels and 735 consonants in the sentence stimuli. Table 4.4 below presents a summary of the metrics calculated. As was mentioned previously, the assumptions of roughly normal distributions did not hold for the data. Therefore, a Kolmogorov-Smirnov test was conducted as it is a non-parametric test that compares the cumulative distributions of two data sets.

Table 4.6 *Summary of metrics for Arabic L1 and English L1 groups visits on consonants and vowels*

Metrics	**English L1**	**Arabic L1**
Sum of all visits on consonants	672	1704
Mean of visits on consonants	21.0	43.7
Sum of all visits on vowels	432	1396
Mean of visits on vowels	13.5	35.8
Mean proportion of visits on vowels	0.38 (006) K-S test=0.85 p= <0.001 t= 4.875 Cohen's d=0.174	0.44 (0.05)

Visits on consonants

As seen in *Figure 4.17* the sum of all the visits on consonants made by the English L1 group was 672, while the sum of all the visits on consonants by the Arabic L1 groups was 1704. The mean number of visits for English L1 was 21.0 while for the Arabic L1, the mean was 43.7.

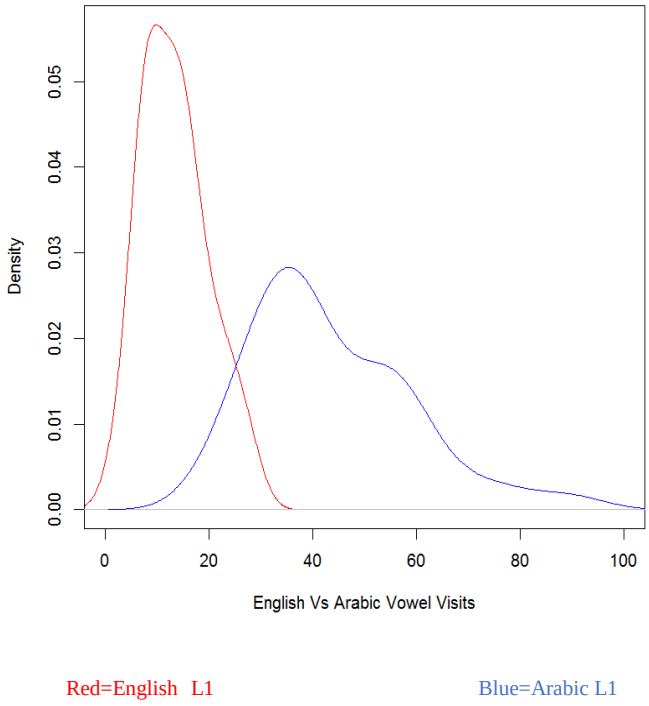


	N	Sum	Mean
L1 English	33	672	21.0
L1 Arabic	39	1704	43.7
Total	72		

Figure 4.17 English L1 and Arabic L1 visits on consonants

Visits on vowels

Figure 4.18 below shows that the sum of all visits on vowels made by the English L1 group was 432, while the sum of the visits on vowels by the Arabic L1 groups was 1396. The mean number of visits for the English L1 participants was 13.5 while the mean number of visits by the Arabic participants was 35.8.

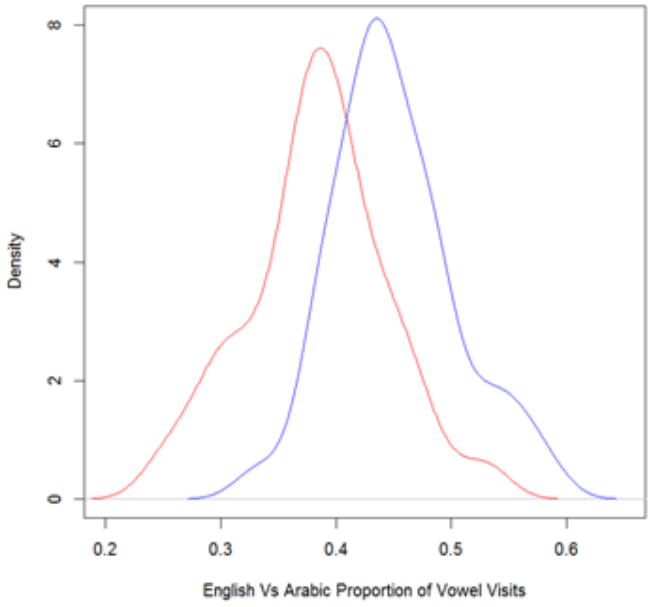


	N	Sum	Mean
L1 English	33	432	13.5
L1 Arabic	39	1396	35.8
Total	72		

Figure 4.18 English L1 and Arabic L1 visits on vowels

Mean Proportion of visits on vowels

Figure 4.19 below shows the mean proportion of visits on vowels. For the English L1 participants, the mean proportion was 0.38 while for the Arabic participants it was 0.44. This result is highly significant (*t=4.875; p= <0.001*).



	N	
L1 English	33	0.38
L1 Arabic	39	0.44
Total	72	

Figure 4.19 English L1 and Arabic L1 proportion of vowel visits

4.2.2.3 Discussion

Number of fixations per individual

The highly significant difference (*W = 48; p = <0.001*) between the individuals in the Arabic L1 group (*M = 450*) and the individuals in the English L1 group (*M =252*) in terms of median number of fixations was expected. The overall number of fixations is believed to be negatively correlated with search efficiency and is indicative of difficulty in interpreting the fixated

information (Holmqvist et al. 2011.) A high number of fixations is generally associated with poor, beginning and dyslexic readers who execute more fixations than proficient adult readers (Rayner, 1998).

Median fixation duration

There was a highly significant difference (*W=91836740; p=<0.001*) between the two groups in terms of median length of fixation duration. "A longer fixation duration is often associated with deeper and more effortful cognitive processing" (Holmqvist et al., 2011, p. 381). Children, dyslexic, poor and beginning readers tend to have longer fixation durations than proficient adult readers. The average fixation duration for skilled L1 English adult readers is between 200 and 250 ms. (Rayner 1998). The median fixation durations exhibited here appear to be quite short for both groups, but this is possibly an artefact of the nature of the task itself which was only at the sentence level. Another explanation might be the length, frequency, and familiarity of the words in the sentences, which are factors known to shorten fixation times (Juhasz & Rayner, 2003; Just & Carpenter, 1980; Rayner, 1997; Sereno & Rayner, 2003). There remains, however, a significant difference in fixation duration between the two groups and providing potential evidence that the Arabic L1 EFL readers are processing English texts differently from proficient English L1 readers.

Saccade length

The median saccade length of the L1 English group is significantly longer (*W=47344944; p=<0.001*) than the Arabic L1 EFL group. This result is in line with the initial hypothesis. Saccades are limited in length by the perceptual span. For left-to-right languages, the perceptual span is asymmetric and extends from not more than 3 or 4 letter spaces to the left of the fixation to approximately 14-15 letter spaces to the right of the fixation (McConkie & Rayner, 1975;

Rayner, Well & Pollatsek, 1980). Previous research showed that the perceptual span for Hebrew, a right–to-left language, is also asymmetric, but larger to the left of the fixation (Pollatsek et al., 1975). However, the perceptual span for Arabic had not been investigated until recently. Jordan et al. (2014) used a gaze-contingent window paradigm in which a region of text was displayed normally around each point of fixation, while text outside this region was obscured. Their findings showed for the first time that a leftward asymmetry in the central perceptual span occurs when Arabic is read and provided a new indication that the perceptual span for alphabetic languages is modified by the direction of the reading. It is plausible that the Arabic L1 EFL participants lacked sufficient experience in reading in English to extend their perceptual span to the right, thus making their saccades significantly shorter than the experienced English L1 participants. Saccade length is also influenced by the length of the fixated word and the word to the right of the fixation (White et al., 2005). It is also clear that the spaces between words are used in targeting where the next saccade will land. Spaces provide information about an upcoming word's length in parafoveal vision and therefore spaces are an important cue used by the eye to calibrate how far to jump with each saccade. However, as was mentioned in section 2.4, the Arabic language contains spaces within, as well as, between words. Hence, the saccadic patterns exhibited by the Arabic L1 participants in this study may reflect a word search strategy transferred from their L1. The short saccadic length displayed by the participants in the Arabic L1 group is also characteristic of dyslexic, poor and beginning readers (Rayner, 1988).

Number of forward saccades per individual

The median number of forward saccades per individual for the English L1 participants was significantly lower ($W = 59;\ p=<0.01$) than for the Arabic L1 EFL participants. This is not surprising because as the English L1 participants made longer forward saccades, fewer were needed to process the words in the sentences. The Arabic L1 EFL participants, on the other hand, needed to execute many more short forward saccades to process the text. Frequent, short saccades are characteristic of poor readers and dyslexics (Hyönä & Olson,1995; Rayner,1998; Rayner et al., 2006).

Length of regressions

There was a significant difference ($W=8061788;\ p=<0.001$) between the median regression length for the English L1 group and the Arabic L1 EFL group. In order to be classified as a regression, the saccade needs to move in the opposite direction of the text, but not necessarily in the opposite direction to the previous saccade. Saccadic regressions occur about approximately 10-15 % of the time, in skilled readers (Rayner, 2009; Dussias, 2010). Regression events occur in different sizes: in-word regressions and between-word regressions. Short regressions are most likely due to oculomotor errors, whereby the eyes overshoot the intended saccade target (Samuels, Rasinski & Hiebert, 2011). However, longer regressions often reflect comprehension failures. One explanation for this result may be the extreme difference in the reading proficiency which is characteristic of the two groups. It has been postulated that better readers have a spatial coding of the text which allows them to resolve anaphora by executing a single long saccade to the antecedent (Murray & Kennedy, 1998). Further analysis of the data would need to be carried out to investigate

the types of regressions (intra-versus inter-word) made by each of the groups before any firm conclusions can be reached.

Proportion of saccades that are regressions

There was no significant difference (W= 805; p= <0.001) between the proportion of regressive movements made by the English L1 and Arabic L1 EFL groups. When children start to read, up to 30% of their fixations are regressions and the number declines up through college-age. Poor readers and dyslexics also make more regressions, relative to good readers comparable in age (Ashby, Rayner & Clifton, 2005). Therefore, this result was unexpected. One possible explanation may be the task itself. Readers were asked to read the two sentences to answer the question "Which sentences describe the sign?" Previous research reported that readers tend to make more regressions to a previously read passage segment for re-inspection when this segment helps answer a subsequent question (Christie & Just, 1976; Kennedy and Murray, 1987; Kennedy et al., 2003). Therefore, the English L1 participants might have been induced to make more regressions than they would normally make.

Mean proportion of visits on vowels

There was a highly significant (t= 4.875; p= ≤0.001) difference between the two groups in mean proportion of visits on vowels. This finding was unexpected, as according to the 'vowel blindness' hypothesis, it was anticipated that the Arabic L1 participants would execute a lower proportion of visits on vowels than the English L1 participants.

Ryan and Meara (1991) suggested that Arabic L1 reading processes lead to a type of 'vowel blindness' when processing English words. Ryan (1997) stated that "Arabic speakers are seriously confused by the excessive amount of information present in English where all the vowels are

written down, a convention which is markedly different from Arabic" (p. 186). She believed that 'vowel blindness' is a condition Arabic L1 speakers experience when reading in English which may be due to a lack of awareness of the function which vowels perform in English. Hayes-Harb (2006) replicated Ryan and Meara's (1991) study and provided evidence that Arabic L1 speakers are "less sensitive" to deleted vowels than either of her control groups. The question arises as to why these studies found evidence for the vowel blindness hypothesis and my study did not. The most obvious reason is that Ryan and Meara (1991) did not have the technology to investigate the eye movements of their participants in real time. In both studies, the task that the participants were asked to complete was a vowel deletion task. However, not being able to recognize deleted vowels is not equivalent to showing a lack of awareness of the vowels (Sadhwani, 2005). This is not a criticism of these studies. They provided valuable information about how Arabic L1 speakers read in English. However, it has been over 25 years since Ryan and Meara's (1991) study and both technology and our understanding of noticing and awareness has broadened and changed.

The more recent study by AlSadoon and Heift (2015) was premised on the assumption that 'vowel blindness' was a proven phenomenon. They reported that the existence of 'vowel blindness' was significantly reduced for the experimental group who received vowel training in the form of textual enhancement. However, as mentioned in Chapter, 3.1.1 there are several limitations to this study in terms of methodology, data analysis and interpretation of findings. Therefore, the results of their study must be interpreted cautiously.

In Study One, although the Arabic L1 EFL participants made more visits on consonants than the English L1 participants, they also made a significantly higher proportion of visits on vowels than the English L1 participants. Consequently, this study provides preliminary evidence that the Arabic L1 EFL students are not unaware of vowels when they read in English. In fact, if

we take measures of fixation duration, number of visits and proportion of visits on vowels, as indications of increased attentional processing (Just & Carpenter, 1980), we can claim that they allocate more attention to the vowels than the English L1 participants. Sadhwani (2005) proposed that it "may not be so much a case of vowel blindness, but rather a matter of vowel confusion and the inability to discriminate between vowels. The problem is a matter of vowel substitution" (p. 6). In her study, significant findings were discovered in participants' word dictation to show that the Arabic L1 speakers did in fact use vowels when writing in English. She suggested that the participants did not experience a visual handicap in terms of vowel function but in terms of vowel choice. Khan (2013) suggested that 'vowel blindness' is only temporary and may be overcome by proper instruction. I would argue that the increased attentional processing of vowels is due to lack of automatic processing, which makes the decoding of the vowels effortful. Therefore, I suggest we discontinue the use of the term 'vowel blindness' as it is not an accurate reflection of how Arabic L1 EFL learners process words while reading in English. Clearly, the Arabic L1 participants were not 'blind' to the vowels, spending significantly more time focussing on them than the English L1 participants.

4.2.2.4 Conclusion

Study One investigated the eye movements of Arabic L1 EFL and English L1 participants while reading in English at the word and sentence level. There were robust differences in the eye movement patterns of the two groups. Results showed that the groups differed significantly in 10 of the 11 metrics calculated. Specifically, the Arabic L1 participants displayed significantly more and longer fixations than the English L1 participants. They also made significantly more and shorter forward saccades than the English L1 group. The regression length of the Arabic L1 group

was also significantly shorter. However, the proportion of saccades that are regressions did not reach significance levels. Much of the variability in these measures is related to the ease or difficulty associated with understanding the text (Rayner, 1998). The data support previous findings that Arabic L1 students have difficulty reading English texts (Abu-Rabia, 1997b; Fender, 2003, 2008; Hayes-Harb, 2006; Randall and Meara, 1988; Randall 2007; Randall & Groom, 2009; Ryan, 1997; Ryan and Meara, 1991; Saigh & Schmitt, 2012; Thompson-Panos and Thomas-Ruzic, 1983). In fact, the reading patterns displayed by the Arabic L1 group with many, long fixations, many short forward saccades and numerous regressions are characteristic of the patterns displayed by beginning and poor readers as well as dyslexics (Ashby, Rayner & Clifton, 2006; Chase, Rayner & Well; Rayner, 1998; Rayner et al., 2006).

It is possible that the differences observed between the Arabic L2 and English L1 readers might be due not only to the different orthographic systems, but also to language proficiency. Unfortunately, due to limitations in access to a participant pool with higher proficiency Arabic L1 readers, this possibility could not be investigated. Students in the department in which I teach range in proficiency from CEFR A1 to CEFR B2.

As mentioned in Chapter 2, Hayes-Harb (2006) concluded that the Arabic L1 EFL students in her study exhibited a pattern of attention to vowels and consonants which differed from English L1 speakers, as observed by response time, accuracy rate and letter detection. She found that the results for vowels versus consonants differed between the Arabic and English groups and between the Arabic and non-Arabic ESL groups but did not differ between the non-Arabic ESL and English groups. Hayes-Harb concluded that because the performance by the Arabic L1 speakers on the letter detection task "differed so markedly from performance by the other groups, the findings cannot simply be attributed to word processing differences between native and non-native readers

of English" (p. 335). Because all the L2 participants in her study were of comparable English proficiency, she postulated that this was further evidence of the exceptional difficulties for native Arabic speakers when processing English. Fender's (2003) study investigated word recognition skills in Arabic L1 and Japanese L1 speakers and found that the Arabic L1 participants were significantly slower and less accurate than the native Japanese speakers. The results from Study One confirm previous findings that Arabic L1 EFL students have difficulties which are particularly characteristic of their language background.

Study One investigated the allotment of visual attention paid vowels and consonants by Arabic L1 and English L1 speakers as they read sentences in English. The finding that the Arabic L1 EFL participants spent more time attending to vowels than did the English L1 participants questions the 'vowel blindness' hypothesis as proposed by Ryan and Meara (1991).

In conclusion, evidence from this study reveals that the eye movements of the Arabic L1 EFL participants and the skilled English L1 participants are qualitatively different as evidenced by the heat maps and gaze plots and quantitatively different, as reported in the statistical analyses for number and duration of fixations, number and length of saccades, regression length and mean proportion of visits on vowels. As mentioned previously in this chapter, contrasting native and non-native reading patterns can potentially generate information as to how 'close' or 'far away' the learner's reading behaviour is to that of a typical skilled reader. This knowledge can then inform pedagogical interventions to help remediate reading difficulties and assist the students to become more efficient and effective readers of English. Chapter 5 describes a study during which such classroom interventions were implemented with the aim of developing the word reading skills and potentially the overall reading comprehension of Arabic L1 EFL students.

Chapter 5. Study Two: An eye-tracking study on the effects of an intensive reading intervention programme with Arabic L1 learners of English

5.1 Introduction

In this chapter, I will describe Study Two and its design. I will give a detailed description of the study which compared two groups of Arabic L1 EFL students, before and after classroom reading interventions, as assessed by eye movements and a pre-and post-reading proficiency instrument. Finally, I will report on the results of both measures and discuss possible explanations for the findings.

5.2. Rationale for the design

Study One showed that the Arabic L1 EFL students paid more attention to vowels than did the English L1 participants. In an attempt to remediate what I considered to be confusion regarding vowels, I created materials which I hoped would make the students more aware of the vowels in the texts and exercises they were reading. The tracking exercises and textual enhancements were applied based on the findings of Study One. The other interventions were based on best practices in the literature regarding the remediation of reading difficulties.

Study Two reconsiders the question of whether input enhancement can make targeted forms more perceptually salient. Several researchers voiced concerns regarding the methodologies employed in input enhancement research (see for example Izumi, 2002; Jahan and Kormos 2014; Winke 2013). The particular design of Study Two addresses these concerns in several respects. The participants had multiple opportunities to experience the treatment conditions, the interventions were carried out over an extended period of time, they were implemented in

classrooms typical of the Middle East context and the instruction was provided by the regular classroom teachers. That is to say, the conditions under which the research was carried out were representative of how teachers use input enhancement with reading goals. Thus, as suggested by Jahan and Kormos (op.cit.), the enhancements were employed along with explicit instruction to foster learners' explicit knowledge of the targeted forms. The present study is in keeping with Winke's (2013) suggestion that any future "input enhancement research be conducted in conditions more representative of how teachers use input enhancement in the real world" (p. 343). The participant sample was representative of Arabic L1 college students in the Gulf Cooperation Council (GCC) countries, in particular those from Qatar, the United Arab Emirates and Saudi Arabia whose education ministries emulated the Egyptian system of education (Al Fadala, 2015). Although there are unique aspects of educational practices in the other GCC countries (Bahrain, Kuwait, and Oman) the commonalities among them are more prominent (Al Fadala op.cit). Therefore, the outcomes of this study should be generalizable to other classrooms of Arabic L1 EFL students in the GCC countries.

Study One showed that the Arabic L1 EFL students displayed eye movements which were similar to poor readers and dyslexic readers. Therefore, in addition to the textual enhancement, I also used interventions which the literature showed to be effective in helping struggling readers and which were explained in Chapter 3. The specific interventions were activities focusing on rapid word recognition and automaticity, phonemic awareness, prosody and spelling (see section 5.5.1.4 for description of how these interventions were implemented).

5.3 Rationale for the in-house materials

Alderson (1984) suggested that if the cause of foreign language reading problems is poor L2 knowledge, then it would make sense in the teaching of a foreign language to concentrate upon teaching and improving this language knowledge. Therefore, reading courses should be more concerned with teaching language competence, rather than with providing reading strategies. Additionally, if the hypothesis is correct that there is a threshold level (Clarke, 1980), beyond which learners have to pass before they can apply reading strategies, then the language competence of struggling readers needs to be raised before the teaching of strategies is effective. For the most part, the text books, used at the technical college in which Study Two took place, focus on relatively higher-order reading strategies, such as scanning for specific information, skimming for gist, understanding vocabulary from context, making inferences and deciphering text structure etc. However, for most learners in the Middle East, the problem is not so much that of understanding at the text level, but of comprehension at the word level (Randall & Meara, 1988). Ryan (1997) proposed that it is precisely at the level of the word that difficulties arise for Arabic L1 EFL students, and therefore context does not always help with comprehension. A further shortcoming of the textbooks used at the college is the absence of fluency exercises for reading. If we acknowledge the importance of word recognition to fluent reading ability, then classroom instruction that focuses on automatic word recognition should be a critical element of any reading programme designed for Arabic L1 EFL students. Because of this lack of emphasis on fluency "in many classes, little actual reading occurs, with the most time devoted to tasks and activities that assume the reading of the text" (Grabe, 2009, p. 379).

All the reading passages/texts used in Study Two were written or adapted for use with the target audience. The topics are culturally sensitive, relevant to the students' lives and tap into the

background knowledge which is crucial for effective reading comprehension. The vocabulary was recycled and attention was paid to presenting grammar systematically with a clear progression. All texts and activities had been piloted in previous semesters and feedback from both instructors and students was incorporated into the final versions.

5.4 Rationale for the medium of instruction and assessment

Paper-based reading and assessment

When designing Study Two, I had to decide whether to use digital or paper-based reading materials. I work in a college which places a strong emphasis on technology and, as in many other educational institutions, books are being challenged by an increasing number of digital devices such as computers, laptops, e-books, tablets and smart phones (Mangen, Walgermo & Brønnick, 2013). There is also a growing trend to promote BYOD (Bring Your Own Device) to classes. However, one of the effects of digital reading is to shift the balance from continuous reading to reading on 'the prowl' (Baron, 2015). During my first years in the Middle East, I noticed that my students 'read' to find answers to questions; they did not read to understand. When questioned, they reported that they had been taught to find a key word in the question and match it with a key word in the text. This confirmed what I had been told by the teachers in a high school where I had conducted some EFL teacher education courses. Students confessed to understanding nothing in the readings themselves, but had managed to pass their previous reading courses in this manner. What the students had learned was to 'hunt and peck' (Baron, op.cit.) to find answers to questions. However, I wanted them to engage in deep reading, which is what Wolf and Barzillai (2009) define as "the array of sophisticated processes that propel comprehension and that include inferential and deductive reasoning, analogical skills, critical analysis, reflection and insight" (p. 32). According to these authors, deep reading may be threatened by today's emphasis on immediacy, and a

cognitive set that embraces speed. Mangen, Walgermo and Brønnick (op.cit.) found that reading linear narrative and expository texts on a computer screen led to poorer reading comprehension than reading the same texts on paper. They suggested that part of the problem with reading continuous text on screen is the difficulty in constructing a mental map of the entire passage. As discussed in Chapter 2.1, the ability to create a mental model is an essential component of the reading process. Disrupted mental maps of the text may be reflected in poorer understanding and ultimately poorer recall of presented material. Cataldo and Oakhill (2000) found that good readers were significantly better than poor readers at remembering and relocating the order of information in a text, which would imply a relationship between mental reconstructions of text structure and reading comprehension. Liu (2005) found that when reading on screen, people tended to browse and scan to look for key words and to read less in a linear fashion. However, on the printed page, they tended to concentrate more on following the text sequentially. In a recent study, Mangen, Velay, Robinet, and Olivier (2014) had students read a short story. When readers were asked to place a series of events in chronological order, those who had read the story in print scored significantly higher, made fewer mistakes and recreated a more accurate version of the story. Ackerman and Goldsmith (2011) concluded that readers perceive the medium of print as more suitable for effortful learning whereas electronic medium is better suited for fast and shallow reading of short texts such as news, email and notes. They suggested that the perception of digital presentation "as an information source intended for shallow messages may reduce the mobilization of cognitive resources that is needed for effective-self regulation" (p. 29).

Although there is still no longitudinal data regarding the effect of digital reading on reading skills development, the results of the aforementioned studies have serious pedagogical implications. We should not presume that there will be no significant impact on reading

performance when we change the format from print to screen, even for shorter texts on reading assessment instruments (Mangen, Walgermo & Brønnick 2013). Consequently, all reading texts and ensuing questions implemented in Study Two were paper-based.

Handwriting

Another choice that had to be made when designing Study Two was whether to use keyboards or handwriting for the intervention activities and assessments. I chose the medium of pen and paper for several reasons. Smoker, Murphy and Rockwell (2009) studied the word retention of 61 American university students. Words were taken from the sixth grade Florida Comprehensive Assessment Test (FCAT). Smoker et al. (op.cit) concluded that because of the additional kinaesthetic information provided by handwriting, participants in their study remembered words better when they wrote them by hand than when they typed them on a keyboard. Longcamp et al. (2008) taught adult English L1 participants how to produce sets of unknown characters either with pen and paper or on a computer keyboard. Participants underwent three training sessions where they learned to write two sets of 10 unknown characters modified from the Bengali and Guajarati alphabets. In each session, each of the 10 characters was written or typed 20 times. Results of character recognition tests showed that correct response rates were higher when the characters had been written by hand than when typewritten. There was a response accuracy decrease over time, with a larger decrease for typed than handwritten characters. Additionally, brain scans showed that when the participants who did their training by hand were asked to remember the characters, the motor function area of their brain became active, which was not the case for those trained on computers. Mangen, Anda, Oxborough and Brønnick (2015) explored the effects of writing modality on word recall and recognition, investigating three writing

conditions: handwriting with a pen and paper, typewriting on a laptop keyboard and using an iPad touchscreen. Participants listened to a series of words that were read aloud to them and they were instructed to write down each word immediately after hearing it. Three word lists were used for this task, each list consisting of 28 semantically related words from three semantic sub-categories: action verbs, animals and food. Results showed that the handwriting condition was associated with better free recall of word lists compared to recall of the word lists written using a keyboard; however, there was no difference between the writing modalities with respect to word recognition. Although the results do not support a simple view that keyboard use per se interrupts memory of written words, the researchers suggest that there may be certain cognitive benefits to hand writing. Mangen, Anda, Oxborough and Brønnick (op.cit) propose that the "graphomotor processes in the handwriting condition may have facilitated a richer encoding of the words into long-term memory, resulting in better retrieval as evidenced in the free recall measure" (p. 240). These studies on character and word recall indicate that there may be cognitive benefits to handwriting which may not be available in keyboard writing.

Baron (2015) argued that a tablet offers a "menu of distraction that can fragment the reading experience, or stop it in its tracks" (p. 8). Mueller and Oppenheimer (2014) suggested that even when laptops are used solely to take notes, they may still be impairing learning because their use results in shallower processing. In three studies, they found that participants who took notes on laptops performed worse on conceptual questions than students who took notes longhand. When students took notes by hand they selected, processed, summarized and reframed information. However, when typing, the faster speed allowed them to transcribe the contents of the lecture verbatim without processing the content. Mueller and Oppenheimer (op.cit) concluded that laptop use can negatively affect performance on educational assessments and warn that laptop use in

classrooms be regarded with caution in spite of its growing popularity. There are situations in which laptops are beneficial for assessment and do not have a negative effect on performance i.e. special access/accommodations for students with learning differences. However, it does seem plausible that the use of laptops might have a negative effect on learning outcomes.

Consequently, all written activities in Study Two were carried out in longhand. Longhand practice appeared to be a logical option as, in addition to the arguments previously put forth in favour of pen and paper use, all students in the study had to take the IELTS the following semester, the writing section of which is a pen and paper essay.

5.5 Introduction to Study Two

Study One found that the Arabic L1 EFL participants displayed eye movement patterns which were significantly different from, and potentially less effective than, the patterns executed by skilled English L1 readers. Study Two was carried out to ascertain if focused pedagogical interventions could help to develop word level reading skills which would be manifested in the Arabic L1 EFL students' eye movements approximating those of skilled readers of English. Study Two also investigated whether these 'teacher engineered enhancements' (Sharwood Smith, 1991) had any significant effect on the overall reading comprehension of the participants. Study Two addressed the following Research Questions:

1) Do focused intervention activities change the eye movement patterns of Arabic L1 EFL students while reading at the word and sentence level in English? If so, how and to what extent?

2) Do focused intervention activities influence word and sentence level reading processes and thus overall reading comprehension? If so, to what extent?

5.5.1 Methodology

5.5.1.1 Participants

Participants were a convenience sample of 39 Arabic L1 EFL students. Participation in the study was voluntary, and sanctioned by the Dean of Language Studies and Academics at the college. All participants read and signed a consent form which explained the general nature of the research and its voluntary participation. They constituted the same group of students whose eye tracking data were collected for Study One. Students were enrolled in five sections of FL 1080, which is a course designed for students at a B1 level of the Common European Framework of Reference (CEFR) (2001).There were 28 females and 11 males. 20 participants were in the 18-20 year range, 15 in the 21-24 range and 4 in the over- 24 range. They were all Qatari nationals who had gone through the same educational system in Qatar. The participants had normal or corrected-to-normal vision. Background information and results for all participants were anonymised. As in Study One, student participation was encouraged by the teachers deducting one low quiz/ assignment from the students' final grade, an incentive which was approved by the Dean of Language Studies and Academics.

Five teachers participated in the study. All five were EFL instructors, each with over 20 years of EFL teaching experience. All had been teaching in the Middle East for at least 10 years. Three were male and two were female. The teachers who taught the two experimental groups were aware of the exact nature of the research, while the teachers of the three control groups were aware only that the research was investigating how Arabic L1 EFL students read in English.

5.5.1.2 *Design*

The study had a quasi-experimental design with pre-treatment and immediate post-treatment eye movement recordings and a pre-test and immediate post-test of overall reading comprehension. Although delayed-post tests would have been desirable, logistics prevented this from being carried out. The study included two intact classes in the experimental group (N=20) and three intact classes in the control group (N=19). The interventions were carried out as part of the standard curriculum in students' regular FL classes. The classes were 17 hours per week over the regular semester which consisted of 12 weeks. Reading instruction, as outlined in the course syllabus, accounted for 40% of the course content and 40% of the final exam grade.

5.5.1.3 Materials

Texts

At the time of this research, FL 1080 was the second of three English for Academic Purposes (EAP) courses designed to prepare students for the School of Business faculty at the college. The in-house materials used in Study Two were developed by looking at the themes of the textbooks in introductory level courses in the School of Business, and either writing new texts or modifying existing ones to make them more linguistically accessible to the students in FL 1080. Using the texts without adapting them was not considered, as the texts were from books used at the Canadian counterpart of the college and were written for an English L1 target audience. The 28 texts used in Study Two were based on the following themes: 1) Recruitment and Retention 2) Managing a Business 3) Information Technology and E-Commerce 4) Marketing Strategies, 5) Conducting Business in a Global Economy 6) Counterfeiting and 7) Employee Compensation. Included in the 28 texts were also several texts on study skills, Qatarisation and Qatari culture.

Text difficulty

The Flesch-Kincaid Grade level metric (Kincaid, Fishbourne, Robert, Rogers & Chissom, 1975) is a measure of text difficulty which is based upon the length of words (the number of letters or syllables) and the length of sentences (number of words). This measure accounts for some of the sources of difficulty in a text. However, it explains only part of text comprehension and ignores many features that are significant in estimating comprehension difficulty. According to Crossley, Greenfield and McNamara (2008), traditional readability indices are "narrowly based on surface-level features and take too little account of the processes a reader brings to the text" (p. 475). They measure only the superficial characteristics of the text which are indicative of a reader's surface understanding i.e. understanding of the words and individual sentences (McNamara, 2013). However, there is now a growing recognition that there are a number of additional factors which contribute to text difficulty (McNamara & Graesser, 2012). Traditional measures are inclined to overlook the importance of readers' deeper level of understanding or "deep reading" (Wolf and Barzillia (2009). Most importantly, for pedagogical purposes, traditional measures of text difficulty do not provide information as to the reasons why a text is difficult and are therefore not helpful to teachers when guidance is needed to diagnose a student's reading difficulties and to plan for remediation (McNamara, Graesser & McCarthy, 2014).

In contrast to the traditional measures of readability, Coh-Metrix is an automated tool that measures cohesion and text difficulty (Graesser, McNamara, Louwerse & Cai, 2004) and runs on the assumption that cohesion is an important aspect of language. Coh-metrix was designed with the goal of improving instruction by providing a means to guide textbook writing and to match textbooks more appropriately to the intended students (Graesser et al., op.cit.). Un-adapted authentic texts are often linguistically inaccessible to EFL students in the GCC, as are many of the

readings in EFL textbooks. Consequently, teachers must adapt much of the material they use in this context. Coh-metrix offered a "more complex picture of the challenges that a reader may face as well as the potential scaffolds that may be offered by the text" (McNamara & Graesser, 2012) and provided important information for the classroom teachers interested in manipulating text difficulty. In fact, Crossley, Allen and McNamara (2011) found that the "variables used in the Coh–Metrix L2 Reading Index are more closely aligned to the intuitive text processing used by L2 material writers when simplifying reading texts than those variables provided by traditional readability formulas" (p. 96).

The eight components measured by Coh-Metrix are:

1) Narrativity-the extent to which words are familiar;

2) Referential cohesion-the extent to which words and ideas in the text overlap across sentences;

3) Syntactic simplicity-the extent to which sentences in the text contain fewer and simpler words and familiar structures;

4) Word concreteness-the degree to which the text contains concrete and meaningful words;

5) Deep cohesion-the degree to which the text contains connectives i.e. causal, intentional and temporal;

6) Verb cohesion–the extent to which there are overlapping verbs in the text;

7) Logical cohesion – the number of explicitly expressed logical relations in the text; and

8) Temporal cohesion – the extent to which the text contains cues about temporality

These eight components are calculated using 103 indices. The Coh-metrix L2 Reading Index is calculated by using three of these indices: a) *word frequency* which uses the CELEX database (Baayan, Piepenbrock & Gulikers 1993), a 17.9 million word corpus b) *syntactic parsing*, which measures the uniformity and consistency of parallel syntactic constructions and c) *word*

overlap, which measures how often content words overlap between two adjacent sentences. A low L2 Readability score indicates that the text will be more difficult for an L2 reader to comprehend, while a high score indicates that the text will be easier.

Table 5.1 below shows the text characteristics of the 28 texts used in Study Two as calculated by Coh-Metrix.

Table 5.1 *Text Characteristics of the 28 Texts*

Title	Coh-Metrix L2 Readability	Number of Words
The Body Shop	18.50	590
Fake Goods Tempting Young Adults	16.38	558
Simply Read	15.61	793
How to Keep Staff Happy	15.46	450
Tips for Language Success	15.26	1164
The World's Richest Arab	14.90	758
Cultural Aspects of International Business	13.82	1101
Recruitment and Retention	12.76	485
Compensating the Workforce	12.16	874
Fake Goods in China	11.90	735
Fake Food	11.71	730
Fake Degrees and HR Management	11.29	955
Qatarisation	11.26	838
Big Blunders in Big Business	11.26	782
Counterfeit Qualifications	10.10	1019
Halal Cosmetics	9.03	551
Wow! What a warranty	8.85	803
Las Vegas Online: Recruitment	8.22	586

Employee training and Development	8.09	671
Counterfeits: From Vans to Vuitton	7.85	732
Fake Cigarettes	7.17	740
Counterfeit Malaria Drugs	7.17	608
Counterfeit Car Parts	6.90	549
Jamba Juice: Recruitment	6.64	609
Fake Drugs	5.79	588
No Cure for Fake Drugs	4.68	599
Interpol: Fake Drugs Swamp the Middle East	4.46	599
Counterfeit Goods: Beyond Fashion	4.10	820

Table 5.2 shows the minimum, maximum and mean word count and minimum, maximum and mean Coh-Metrix readability indices

Table 5.2 *Minimum, maximum and mean word count and Coh-Metrix readability indices*

	Word Count	Coh-Metrix L2 Readability
Min	450	2.45
Max	1019	18.50
Mean	723.89	10.15
SD	180.39	4.28

Tasks

Each text was immediately followed by a set of comprehension questions:

1) five Main Idea questions in which participants were asked to choose from three options;

2) five Vocabulary in Context questions which required participants to select one of three options;

3) five Pronoun Reference questions in which the pronouns were underlined in the text and participants were asked to write the referent in the space provided;

4) five True/False/Not Given questions which required participants to circle either T, F or NG;

5) five Multiple Choice detail questions which required selecting 1 of 3 options;

6) five Short Answer questions which required a written response of from 5 to 10 words;

These question types were mandated by the Assessment Department of the College for all midterm and final exams. Therefore, the in-class reading activities followed the same format so as not to disadvantage students in their formal assessments.

5.5.1.4 Interventions

5.5.1.4.1 Interventions to promote noticing and awareness of vowels.

1) Textual enhancement

In the 28 reading texts given to the experimental groups, the vowels were highlighted in bold and red. An exploratory study carried out in the previous semester investigating the different combinations of font type, font size and colour in reading passages revealed that student preference was red and bold. The font used was Ariel, as the British Dyslexia Association deemed it the most 'reader friendly' (see Appendix C for an example of a reading text).

2) Tracking exercises

The tracking activities used in this study are activities in which the students are given a stimulus word at the beginning of a horizontal line of words. Six words follow across the line and differ from the stimulus word by only one letter, which is a vowel (**bat**: bet, but, bit, bat, bit, bat). Students must underline or circle the words which are the same as the stimulus word (see Appendix D for examples). In addition to encouraging students to pay attention to the vowels in each word, the activities also served to reinforce the left-to-right directionality while reading in English. The exercises increased in complexity from single words to phrases of three to four words as the semester progressed. This type of activity was deemed important as 19% of the words in the NGSL 1,000K (Browne et al., 2013) and 22% in the 2,000K differ only by vowels (McCray, 2016, personal communication).

5.5.1.4.2 Interventions designed to increase rapid word recognition and automaticity

Many of the Arabic L1 EFL students whom I have taught in the Gulf have developed the habit of looking at the first letter of a word, and perhaps the middle and end and then unsuccessfully predicting the rest of the word. This is clearly evidenced when they are reading aloud. A few examples are 'pineapples' read as 'Philippines,' 'century' read as 'country', 'murder' as 'mother', 'exciting' read as 'exploring' and 'participle', as 'particle'. As mentioned in Section 4.2, participants were shown their animated eye-movement recordings after the data were collected in Study One. The recordings of the Arabic L1 participants displayed a "hopping" behaviour with eye movements dipping in and out of a word before proceeding on to the next. The Arabic L1 participants were fascinated by this, especially when they were shown recordings of the English L1 participants where the 'hopping' behaviour was not displayed. I good-humouredly referred to this phenomenon of hopping over letters as 'rabbit reading'. In order to eradicate this behaviour, the following activities were carried out in the experimental groups.

3) NGSL lists

Lists compiled from the first 1,500 words of the New General Service List (NGSL) (Browne, Culligan & Phillips, 2013) were created and given to participants (see Appendix E for an example). As Qian (1999) found that students had a higher rate of retention for decontextualized word lists than for contextualized ones, participants received isolated words in list form. These lists were used as the basis for the Spelling and Rapid Word Reading component implemented with the experimental groups and for the Word Meaning component with the control groups.

4) Rapid Word Recognition (RWR)

Recent research showed that the size of children's sight vocabularies and their word recognition speed had a strong relationship to contextual reading (Morris, Trathan, et. al, 2012; Torgesen, Rashotte & Alexander, 2011). Therefore, a Rapid Word Recognition component was included in the interventions. These activities entailed students taking three columns of 35 words from the NGSL word list and reading them aloud and timing themselves on their mobile phones or watches.

5) Spelling with the NGSL List

Each week, the teachers of the experimental groups would take one page from the NGSL List. On Sundays, the teachers would read aloud the words in the first column on the page. They would demonstrate the syllabification and word stress as well as giving spelling rules as applied to specific words in the column (for example hard and soft 'c', hard and soft 'g', words with a final silent 'e' etc. Students would then read one word each going around the class. They would be told

to study the words for homework. The following day there would be a spelling quiz. The same procedure would be followed on Monday, Tuesday and Wednesday. On Thursday, there would be a quiz on 25 words from the columns chosen by the teacher. There was no instruction on word meanings as there was in the control group.

5.5.1.4.3 Activities designed to encourage fluent reading/prosody/phonological awareness

1) Oral text reading fluency

a) Pair reading. Pair reading is a research-based strategy used with readers who lack reading fluency. In this strategy, students read aloud to each other. When using partners, more fluent readers can be paired with less fluent readers, or students who read at the same level can be paired to reread a text they have already read. Paired reading can be done by taking turns reading a paragraph, a page or longer passage.

b) Relay reading. Relay reading is a type of reading activity in which one student reads one sentence and then the student next to him or her reads the next sentence going around the class. It has been my experience that this type of reading keeps students focussed and on task, reduces mind wandering and forces the students to pay attention to punctuation, especially periods at the ends of sentences.

c) Teacher reading. Word recognition can be improved by the teacher reading aloud while the students follow along with the text. This technique forces the students to recognize the words quickly in order to keep up with the reading and improves the recognition process by providing dual modality: visual and aural (Randall, 2007). It also provides an essential model for phrasing/expression, stress/intonation, rhythm and rate.

For each of the texts, the teachers first read the text aloud with the students listening and following the words on their papers. Then a relay reading activity would follow. As students should be given opportunities to re-read texts (Rasinski, 2003; Stahl & Kuhn, 2002; The National Reading Panel, 2003) re-reads occurred at the end of each week with students using the pair reading technique.

2) Phonological/phonemic awareness

Phonological awareness is the ability to recognize that words are made up of a variety of sound units while phonemic awareness is the understanding of the ways that sounds function in words at the phoneme level (Ehri, 1984; National Reading Panel, 2000). Activities which are commonly used to teach and improve phonological and phonemic awareness are a) phoneme identification b) phoneme count c) phoneme deletion d) syllable identification e) syllable deletion and f) rhyme.

Phonemic awareness instruction did not have a specific time allotted but arose naturally out of daily the spelling, reading and writing activities (see Appendix F for examples of phonemic awareness activities).

5.5.1.5 Assessment of reading proficiency

The reading pre-test consisted of a reading text of 750 words on the topic of counterfeit goods being sold in Qatar. It had a Coh-Metrix L2 readability of 7.67, contained 35 sentences, with a mean sentence length of 21.4 and was 6 paragraphs in length. There were five Main Idea questions, five Vocabulary in Context questions, five Pronoun Reference questions, five True/False/ Not Given questions, five Multiple Choice questions and five Short Answer questions which required fewer than ten words to complete. All items were scored as either right

or wrong (1 or 0). The test had originally been created as an-end-of-semester achievement test and it, along with the answer key, had been administered in two previous semesters with five classes of FL 1080 students (N= 51). All items with Facility Values of below 20 per cent and above 80 percent were investigated and the values interpreted together with the Discrimination Indices. A corrected item-total correlation (CITC) index of +.3 is commonly accepted as showing that an item is discriminating positively (Green, 2013). Looking at the Facility Values and the Discrimination Indices together, judgement calls were made as to the rewording or discarding of any item. Green (op.cit.) noted that we might expect higher Facility Values in an achievement test than in a proficiency test and a lower (CITC) as all test takers are performing well because they have mastered what had been taught. The same assessment was given to participants as a post-test.

Eye tracking Measures

Data analysis was carried out to ascertain if there were any statistically significant differences between the pre-treatment eye movement recordings and the post-treatment recordings. The metrics used in Study One were again used for Study Two: 1) Number of fixations 2) Fixation duration 3) Forward saccade length 4) Number of forward saccades 5) Regression Length 6) Total number of regressions 7) Sum of all visits on consonants 8) Mean of all visits on consonants 9) Sum of all visits on vowels 10) Mean of all visits on vowels and 11) Proportion of visits on vowels. The post-treatment data were collected using the same task as Study One.

5.5.1.6 *Research design of experimental and control groups*

Figure 5.1 below shows the research design of Study Two for the experimental and control groups

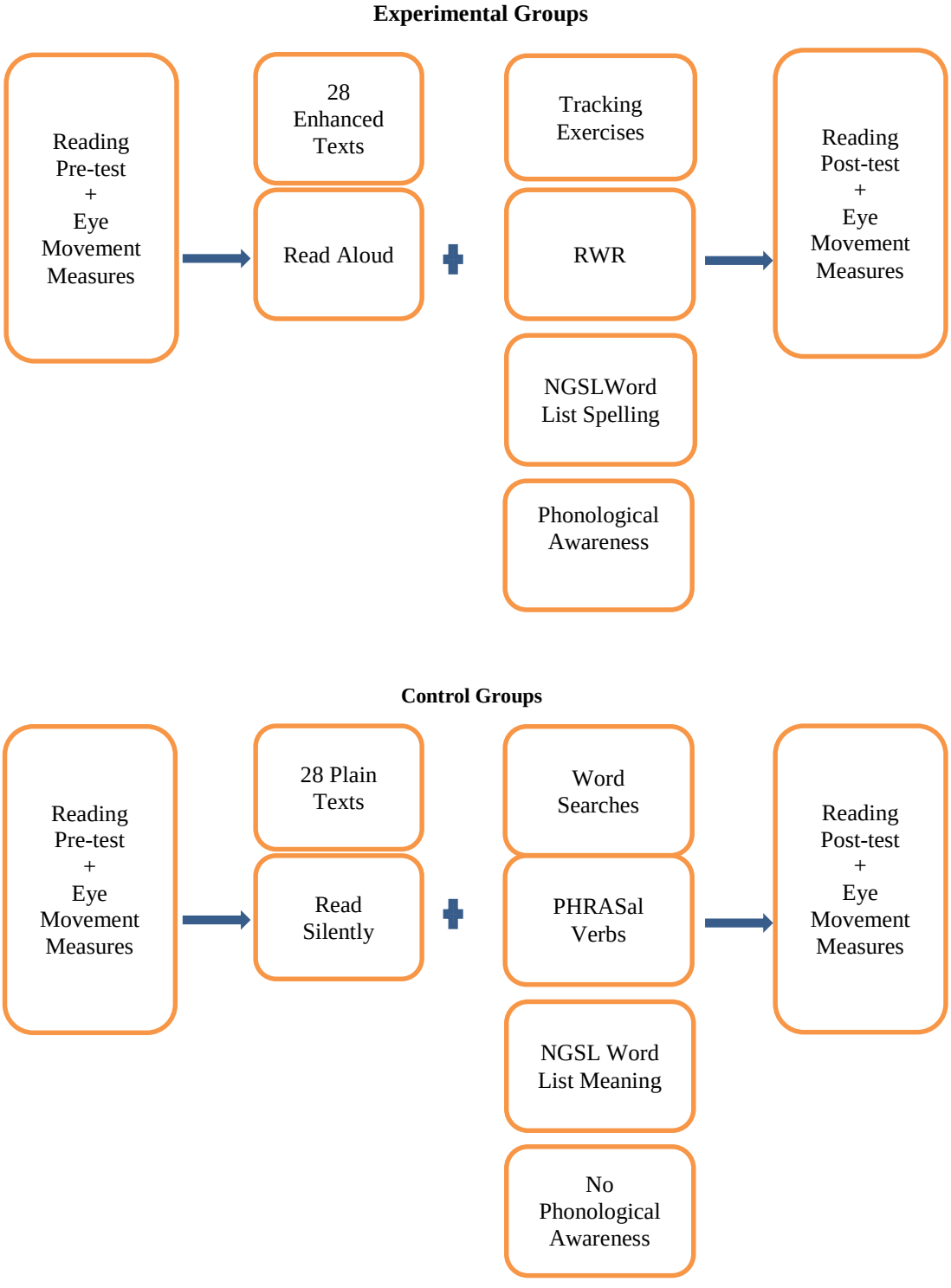


Figure 5.1 Research design for Study Two

5.5.1.7 Procedure

Reading Pre-test

The reading pre-test was administered during the first week of classes before any reading instruction had taken place. Tests were administered by the classroom teachers on the same day during class time. Participants were given one hour to complete the test. All tests were corrected by the researcher and information on individual student scores was given to the classroom teachers for diagnostic purposes.

Reading post-test test

As previously mentioned, the same assessment was given to participants as a post-test under identical conditions during the second to last week of the semester. Participants were not given any feedback or results from the pre-test and as the time lapse between administrations was 12 weeks, there should have been little or no carryover of correct response bias or priming. It was therefore decided not to attempt to create a parallel version of the reading test. Designing another test which would be equivalent in the level of difficulty and topic of the reading passage, the vocabulary used, the length of sentences and the types of questions would have been challenging and might have posed an 'instrumental threat' to validity. Therefore, an argument can be made for using the same version of the test for both pre-and post-assessments.

This was an intervention programme carried out in a regular EFL course where there could be no rigid decisions made as to the exact time taken for each activity. Additionally, the interventions had to be implemented into the regular curriculum in which there were quizzes, tests, exams and assignments mandated by the assessment department of the college. Consequently, it

was impossible to implement the interventions on a regular, rigid schedule. However, Table 5. 3 below shows the approximate number and time given to each of the interventions.

Table 5.3 *Approximate number and time devoted to each intervention*

Interventions	**Approx. number per week**	**Approx. time per week**
Enhanced texts read orally	5 texts	5 hours
Spelling	3 columns of 15 words each	20 minutes
Rapid word recognition	45 words a day x 5 = 225	10 minutes x 5 = 50 minutes
Tracking exercises	2-5 sheets	60 minutes
Phonemic awareness activities	daily but not not scheduled	50 minutes

Eye movement recordings

Data from the Arabic L1 EFL participants in Study One were used as baseline data. Eye movement data were collected again during the second to last week of the semester under the same conditions as in Study One. The metrics used in Study One were again used in Study Two.

Textual Enhancement

Participants in all groups were given the same 28 texts in the same order during the semester. However, the experimental groups received the textually enhanced versions which had all vowels printed in red and in bold (see Appendix G for an example). The control groups were given texts typed in customary black with no bolding. The questions following the texts were

identical for both groups. All texts and questions for all groups were read and completed during class time.

Tracking activities

Participants in the experimental groups were given 20 tracking activities over the course of the semester. These exercises took from 5 to 8 minutes and were usually completed at the beginning of class and served as a warm-up activity before actual reading instruction took place. Participants in the control group completed warm-up activities consisting of word searches using words from the NGSL lists, but they did not do the tracking activities. These word searches contained words displayed only in a left-to-right direction and did not contain words displayed vertically, diagonally or in a right-to-left direction as is common in most word search puzzles. This was to ensure that the experimental groups were not potentially disadvantaged because of the directionality of the words given to them.

NGSL lists

Both the experimental and control groups were given words from the first 1,200 words in the NGSL list at the beginning of each week. As these lists comprised very short and frequent words, one page of 100 words was given each week. Participants in the experimental groups were given a spelling quiz each week from the 100 words. The quizzes with 25-30 of these words chosen by the researcher were given at the end of each week. The teachers would first say a word aloud, then say the word again in a sentence for context, and finally they would say the word again as an isolated word. Participants wrote the words with pen on paper. The control groups were given the same NGSL list as the experimental groups each week but were given activities with meaning-focused instruction without a spelling component (see Appendix H for an example).

Rapid word recognition

All participants in the experimental groups engaged in rapid word recognition activities. The same word list from the NGSL was used. These activities were completed during class time at the beginning, middle and end of each week to measure any improvement in recognition rate. The activities entailed students taking three columns of 15 words and reading them aloud. Participants worked with a partner or a small group to ensure that words were pronounced correctly. Times were recorded by each pair or group and recorded in a log to chart progress in recognition rate. Participants in the control groups were given activities and exercises in the PHRASal Expressions List of 505 of the most frequent non-transparent multiword expressions in English intended especially for receptive use (Martinez & Schmidt 2012) (see Appendix I for an example).

Oral text reading fluency

As mentioned previously, pair reading, relay reading and teacher reading activities were used for the 28 texts read in the experimental groups. The control groups read the same 28 texts but participants read individually and silently as was the regular practice at the FL 1080 level at the college.

5.*6 Analysis*

5.6.1 Pre-and immediate post-test reading results

Tables 5.4 and 5.5 below present a summary of the pre-and immediate post-test reading scores calculated for the Arabic L1 EFL participants in the experimental and control groups. The effect of the treatment on test scores was calculated using a 2-way repeated-measures ANOVA

(Field, 2013). This analysis permitted the investigation of three aspects of the data simultaneously controlling for each. These aspects are 1) the overall differences between the experimental group and control group in terms of score, 2) overall differences in score from Time one to Time two and 3) the interaction between *group* and *time*. This interaction is the important measure which allowed me to judge whether the classroom interventions had a statistically significant effect on the experimental group's test scores. An analysis on the overall test scores for all the items was first performed. Then an analysis was carried out by item type. One problem with these analyses was the fact that the assumption of normality was not met in all cases. Some of the variables were approximately normally distributed and some were not. However, given the fact that this was a relatively small, non-representative convenience sample and that there are multiple comparisons on the same data set, an untransformed ANOVA was appropriate, given the nature and quantity of the data. ANOVA is fairly robust in terms of violations of the assumptions of homogeneity of variance when sample sizes are equal (Field, 2011). The within-subject variable was *time*, (pre- and post-tests), and the between-subject variable was the *group* (treatment or control). After visual inspection of the histogram, no obvious outliers were detected, commensurate with the sample size.

Descriptive statistics for the pre-and post-test reading scores were calculated for the control and experimental groups at Time 1 and Time 2 (see Table 5.2 and 5.3 below).

Table 5.4 *Descriptive statistics for pre-and post-test reading scores at Time 1*

Feature	**Group**	**Min**	**Max**	**Mean**	**SD**	**Skewness**	**SE**	**Kurtosis**	**SE**
Total	Control	7.00	24.00	11.79	4.00	1.49	0.47	2.67	0.92
	Experimental	7.00	19.00	12.21	3.57	0.55	0.52	-0.44	1.01
MI*	Control	0.00	4.00	1.88	0.95	-0.07	0.47	0.39	0.92
	Experimental	0.00	4.00	2.11	1.15	-0.23	0.52	-0.34	1.01
VC	Control	0.00	4.00	1.96	1.08	0.09	0.47	-0.35	0.92
	Experimental	0.00	5.00	2.21	1.18	0.67	0.52	0.88	1.01
PR	Control	0.00	5.00	2.50	1.29	0.00	0.47	-0.76	0.92
	Experimental	1.00	4.00	2.21	1.18	0.45	0.52	-1.30	1.01
RS	Control	1.00	12.00	5.45	2.41	.50	0.47	1.25	0.92
	Experimental	2.00	12.00	5.68	2.60	.87	0.47	0.54	1.01

*MI-Main Idea, VC-Vocabulary in Context, PR-Pronoun Reference. RS-reading for specific information

Table 5.5 *Descriptive statistics for pre-and post-test reading scores at Time 2*

Feature	**Group**	**Min**	**Max**	**Mean**	**SD**	**Skewness**	**SE**	**Kurtosis**	**SE**
Total	Control	2.00	22.00	13.21	3.96	-0.42	0.4	2.41	0.9
	Experimental	11.0	20.00	14.58	2.97	0.20	0.5	-1.25	1.0
MI*	Control	0.00	4.00	2.08	1.21	-0.01	0.4	-1.01	0.9
	Experimental	1.00	4.00	3.00	0.94	-0.44	0.5	-0.82	1.0
VC	Control	0.00	5.00	2.08	1.41	0.14	0.4	-0.63	0.9
	Experimental	1.00	5.00	2.79	2.79	0.17	0.5	-0.47	1.0
PR	Control	0.00	5.00	3.00	1.38	-0.22	0.4	-0.53	0.9
	Experimental	1.00	5.00	3.05	1.03	0.52	0.5	0.37	1.0
RS	Control	1.00	11.00	6.04	2.23	0.50	0.5	0.67	0.9
	Experimental	3.00	11.00	5.73	2.25	0.52	0.5	0.76	1.0

*MI-Main Idea, VC-Vocabulary in Context, PR-Pronoun Reference, RS-Reading for specific information

Table 5.6 *A comparison for Time, Group and Time/Group interaction*

Feature	*Time*	*Group*	*Group/Time*
Total scores	F (1,9.21) p = 0.004	F (1,0.90) p = 0.350	F (1,0.58) p = 0.450
Main idea	F (1,7.17) p = 0.011	F (1,4.98) p = 0.031	F (1,2.778) p = 0.103)
Vocab. in context	F (1,2.72) p = 0.107	F (1,2.51) p = 0.121	F(1,1.13) p = 0.294
Pronoun reference	F (1,13.38) p = 0.001	F (1,0.30) p = 0.725	F (1,0.869) p = 0.357
Reading for specific information	F (1,0.47) p = 0.493	F (1,0.05) p = 0.945	F (1,0.33) p = 0.567

Analysis of total scores

There was a significant main effect with relation to the within-subject variable, *time. F* (1, 9 =.21, p = 0.004. In other words, the participants in both groups performed significantly better on Time 2 than they did on Time 1. However, there was no statistically significant main effect with relation to the between-subject variable, *group F* (1, 0) = .90, p = 0.350. That is to say, the scores, overall, were not significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 0) = .58, p = 0.450 (see Figure 5.1 below). Although there was an increase in score between Time 1 and 2 for the experimental group, that increase was not significantly greater than the increase in score of the control group (see Figure 5. 2 below).

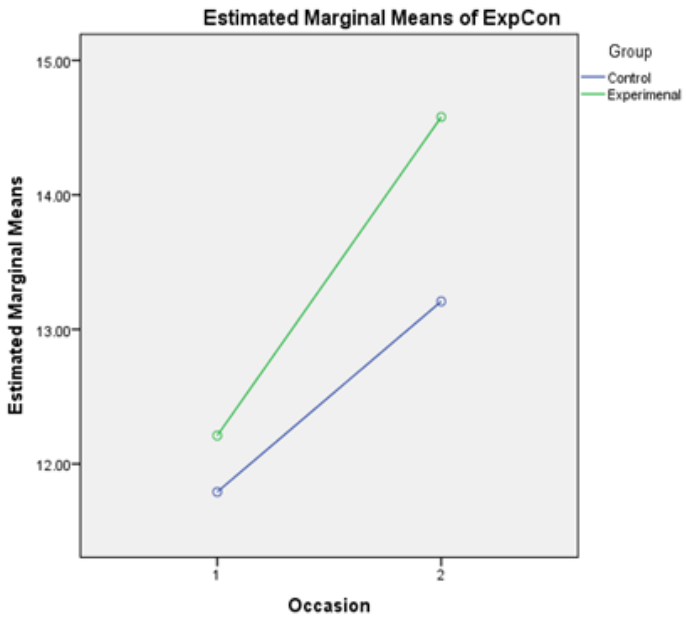


Figure 5.2 Analysis of total scores

Analysis of main idea scores

There was a significant main effect with relation to the within-subject variable, *time*. *F* (1, 7) = .17, *p* = 0.011. That is to say, there was a statistically significant improvement in all the participants' test scores from Time 1 to Time 2. There was also a statistically significant main effect in relation to the between-subject variable, *group*. *F* (1, 4) = .98, *p* = 0.031. There was no statistically significant interaction effect between *group* and *time F* (1, 2) = .778, *p* = 0.103. While there was an increase in scores between Time 1 and Time 2 for the experimental group, that increase was not significantly greater than the increase in score of the control group (see *Figure 5.3* below).

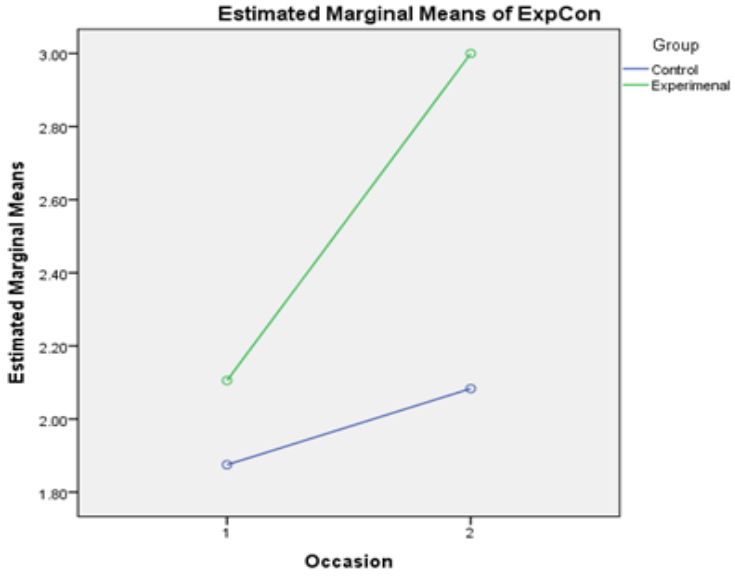


Figure 5.3 Analysis of Main Idea scores

Analysis of Vocabulary in context scores

There was no significant main effect with relation to the within-subject variable, *time F* (1, 2) = .72), *p* = 0.107. The participants did not perform significantly better on Time 2 than they did at Time 1. There was no statistically

significant main effect in relation to the between-subject variable, *group F* (1, 2) = .51, *p* = 0.121. The scores, overall, were not significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 1) =.13), *p* = 0.294. While there was an increase in scores between Time 1 and Time 2 for the experimental group, that increase was not significantly greater than the increase in scores of the control group (see *Figure 5.4* below).

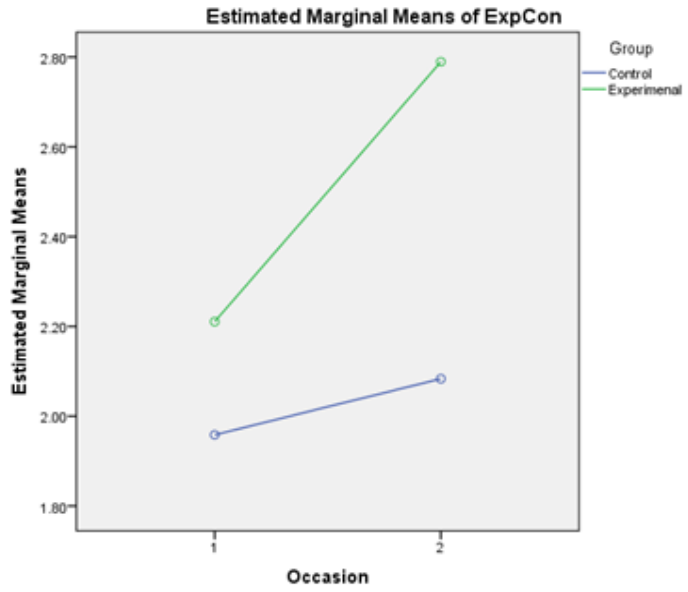


Figure 5.4 Analysis of Vocabulary in Context items

Analysis of pronoun reference scores

There was a significant main effect with relation to the within-subject variable, *time F* (1, 13) = .38), $p = 0.001$. In other words, the participants performed significantly better at Time 2 than they did at Time 1. There was no significant main effect in relation to the between-subject variable, *group F* (1, 0) = .30, $p = 0.725$. There was no statistically significant interaction effect between *group* and *time F* (1, 0) = .869, $p = 0.357$. While there was an increase in scores between Time 1 and 2 for the experimental group, that increase was not significantly greater than the increase in scores of the control group (see *Figure 5*.5 below).

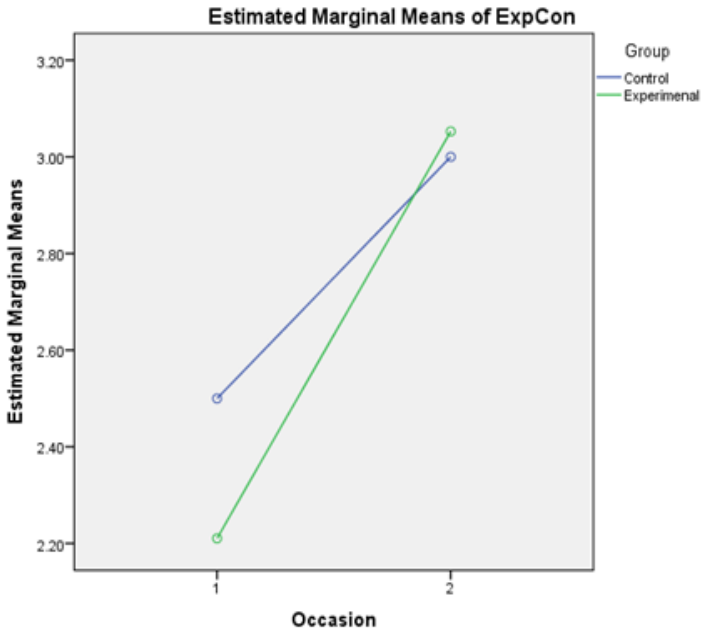


Figure 5.5 Analysis of Pronoun Reference items

Analysis of Reading for specific information scores

There was no significant main effect with relation to the within-subject variable, *time* F (1, 1) = .479, p = 0.493. In other words, the participants did not perform significantly better on Time 2 than they did on Time 1. There was no statistically significant main effect in relation to the between-subject variable, *group* F (1, 0) = .005, p = 0.945. In other words, the scores, overall, were not statistically significantly different across groups. There was no statistically significant interaction effect between *group* and *time* F (1, 4) = .334, p = 0.567.

5.6.2 *Analysis of pre-and post-treatment eye movement metrics*

Tables 5.7 and 5.8 below present a summary of the descriptive statistics for Time 1 and Time 2. Table 5.7 presents a comparison for Time, Groups and Time/Group interaction. An explanation of the results follows.

Table 5.7 *Descriptive statistics for eye movements at Time* 1

Feature	**Group**	**Min**	**Max**	**Mean**	**sd**	**Skew**	**SE**	**Kurt**	**SE**
NFix	Control	341.00	980.00	473.21	154.60	3.05	0.60	10.40	1.15
	Experimental	88.00	703.00	460.50	146.10	-0.67	0.54	1.31	1.04
NSac	Control	192.00	596.00	300.00	93.36	2.68	0.60	8.79	1.15
	Experimental	88.00	450.00	298.22	98.25	-0.34	0.54	-0.06	1.04
NReg	Control	118.00	384.00	172.57	64.95	3.00	0.60	10.01	1.15
	Experimental	88.00	307.00	166.67	50.24	1.30	0.54	2.52	1.04
PropReg	Control	0.26	0.73	0.41	0.11	2.32	0.60	7.65	1.15
	Experimental	0.26	0.50	0.37	0.07	0.33	0.54	-0.72	1.04
MedFix	Control	57.00	159.00	119.54	31.96	-0.66	0.60	-0.39	1.15
	Experimental	34.50	191.00	112.67	37.05	-0.12	0.54	0.41	1.04
MedSac	Control	111.00	277.00	210.71	37.53	-1.22	0.60	3.65	1.15
	Experimental	120.50	325.50	206.81	44.70	0.71	0.54	2.26	1.04
MedRegSac	Control	-209.00	-100.00	-152.07	36.73	-0.01	0.60	-1.14	1.15
	Experimental	-225.00	-81.00	-160.61	34.82	0.81	0.54	1.52	1.04

Table 5.8 *Descriptive statistics for eye movements at Time 2*

Feature	**Group**	**Min**	**Max**	**Mean**	**sd**	**Skew**	**SE**	**Kurt**	**SE**
NFix	Control	327.00	526.00	420.29	60.62	0.39	0.60	-0.49	1.15
	Experimental	235.00	607.00	431.89	120.5	-0.27	0.54	-1.02	1.04
NSac	Control	224.00	335.00	270.00	35.42	0.26	0.60	-0.81	1.15
	Experimental	120.00	431.00	281.39	91.21	-0.22	0.54	-0.63	1.04
NReg	Control	103.00	216.00	149.93	35.11	0.91	0.60	-0.11	1.15
	Experimental	91.00	274.00	150.50	45.27	1.18	0.54	1.97	1.04
PropReg	Control	0.30	0.46	0.35	0.05	0.93	0.60	0.80	1.15
	Experimental	0.25	0.49	0.36	0.07	0.50	0.54	-0.62	1.04
MedFix	Control	81.50	173.00	128.64	39.96	-0.14	0.60	-1.61	1.15
	Experimental	2.50	171.50	118.00	42.53	-1.05	0.54	1.74	1.04
MedSac	Control	161.00	242.50	212.71	21.86	-1.08	0.60	1.12	1.15
	Experimental	140.00	277.00	208.64	42.74	-0.15	0.54	-0.88	1.04
MedRegSac	Control	-231.00	-85.00	-148.75	39.88	-0.43	0.60	0.03	1.15
	Experimental	-224.00	-78.50	-152.67	38.79	-0.10	0.54	-0.34	1.04

Table 5.9 *A comparison for Time, Group and Time/Group interaction*

Feature	Time	Group	Group/Time
NFix	F (1,3.56) p =0.069	F (1,0.00) p = 0.989	F (1,0.32) p = 0.578
NSac	F (1,2.85) p = 0.102	F (1,0.03) p = 0.860	F (1,0.23) p = 0.639
NReg	F (1,6.42) p = 0.017	F (1,0.03) p = 0.869	F (1,0.18) p = 0.675
PropReg	F (1,4.25) p = 0.048	F (1,0.78) p = 0.385	F (1,1.73) p = 0.189
MedFix	F (1,1.70) p = 0.203	F (1,0.55) p = 0.464	F (1,0.12) p = 0.736
MedSac	F (1,0.25) p = 0.625	F (1,0.09) p = 0.765	F (1,0.00) p = 0.983
MedSacReg	F (1,1.62) p = 0.212	F (1,0.24) p = 0.625	F (1,0.27) p = 0.605

Median number of fixations

There was no significant main effect in relation to the within-subject variable *time F* (1, 3) = .56), $p = 0.069$. The participants did not make significantly fewer fixations at Time 2 than on Time 1. There was no statistically significant main effect in relation to the between-subject variable *group F* (1, 0) = .00, $p = 0.989$. In other words, the mean number of fixations, overall, was not significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 0) = .32, $p = 0.578$. While there was a decrease in mean number of fixations between Time 1 and 2 for the experimental group, this decrease was not significantly greater than the decrease in mean number of fixations of the control group. *Figures 5.6* and *5.7* below shows the mean number of fixations at Time 1 and 2 for the Experimental and Control groups.

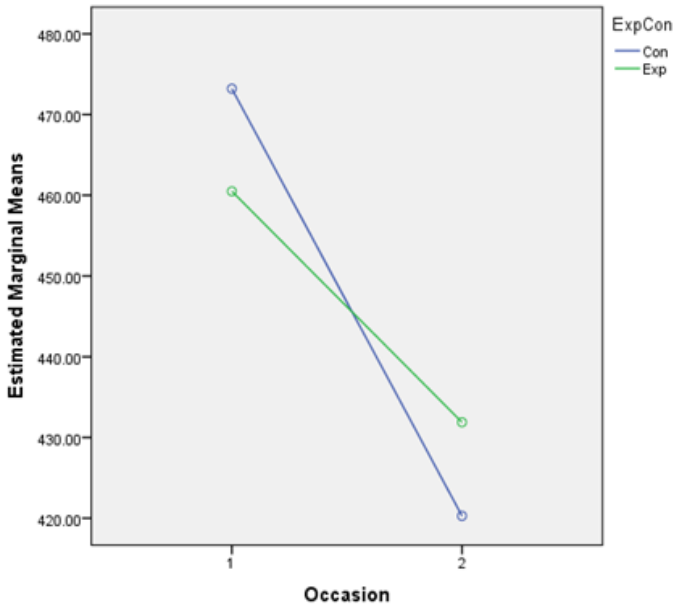


Figure 5.6 Median fixation plot

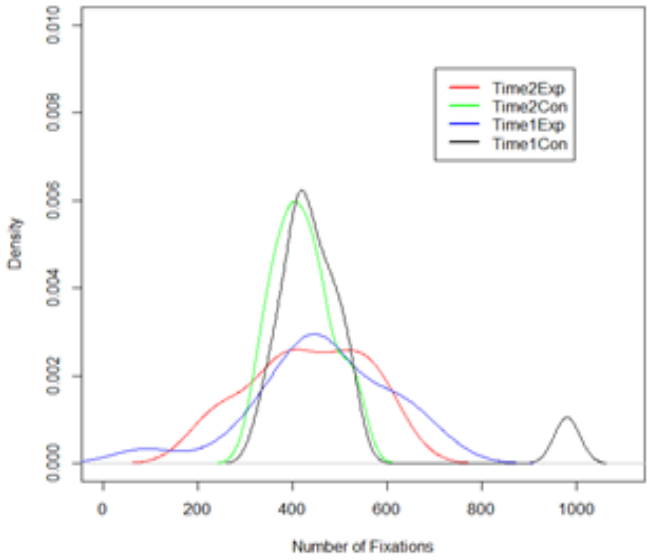


Figure 5.7 Median fixation density plot

Median fixation duration

There was no significant main effect in relation to the within-subject variable *time* F (1, 1) = .70, p = 0.203. The median fixation duration was not significantly different at Time 2 than at Time 1 for both groups. There was no statistically significant main effect in relation to the between-subject variable *group* F (1, 0) = .55, p = 0.464. In other words, the median fixation duration, overall, was not statistically significantly different across groups. There was no statistically significant interaction effect for the interaction between *group* and *time* F (1, 0) =.12, p = 0.736. While there was a decrease in fixation duration between Time 1 and 2 for the experimental group, that decrease was not significantly greater than the decrease in fixation duration of the control group.

Median saccade length

There was no significant main effect in relation to the within-subject variable, *time F* (1, 0) = .25, *p* = 0.625. In other words, the median saccade length was not significantly longer at Time 2 than at Time 1. There was no statistically significant main effect in relation to the between-subject variable *group F* (1, 0) = .09, *p* = 0.765. The median saccade length, overall, was not statistically significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 0) = .00, *p* = 0.983. In other words, although there was an increase in median saccade length between Time 1 and Time 2 for the experimental group, that increase was not significantly greater than the increase for the control group.

Number of forward saccades

There was no significant main effect in relation to the within-subject variable *time F* (1, 2) = .85, *p* = 0.102. The number of forward saccades was not significantly different from Time 1 to Time 2. There was no statistically significant main effect in relation to the between-subject variable *group F* (1, 0) = .03, *p* = 0.860. In other words, the number of forward saccades, overall, was not statistically significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 0) =.23, *p* = 0.639. In other words, while there was a decrease in the number of forward saccades between Time 1 and 2 for the experimental group, that decrease was not statistically significantly different from the decrease in number of forward saccades of the control group. *Figures 8* and *9* below show the number of forward saccades for the experimental group and control group at Time 1 and Time 2.

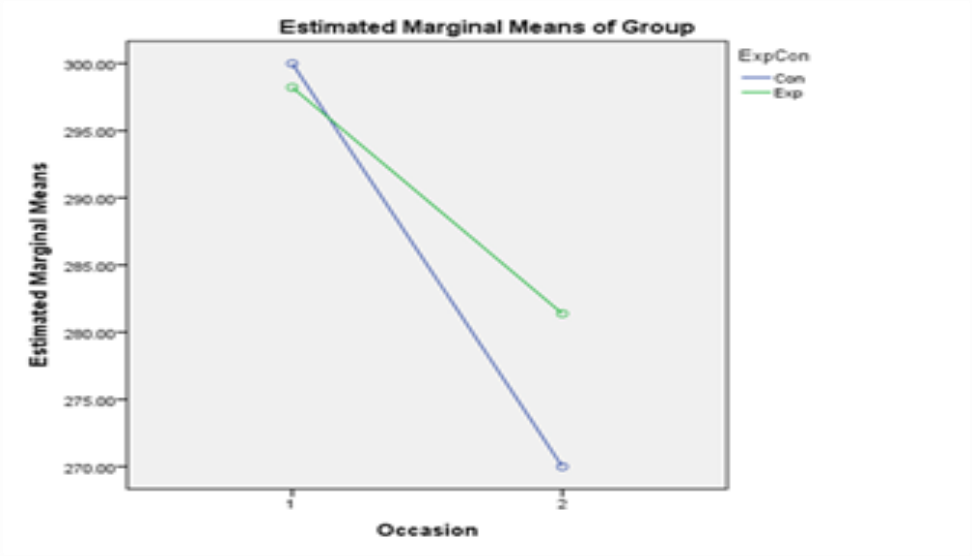


Figure 5.8 Number of forward saccades plot

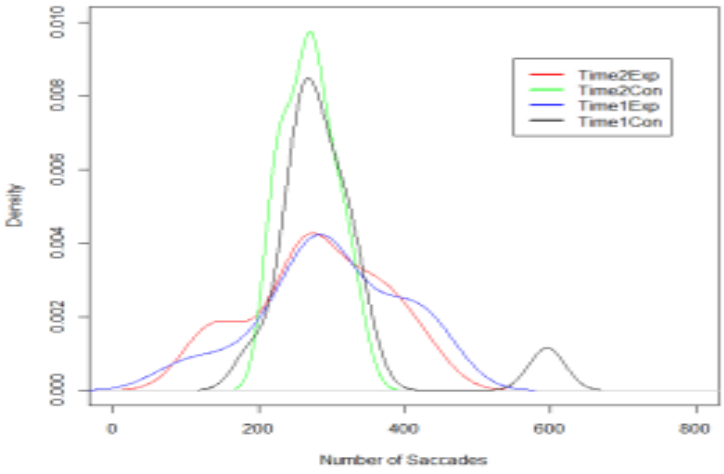


Figure 5.9 Number of forward saccades density plot

Median Regression length

There was no significant main effect in relation to the within-subject variable *time F* (1, 1) = .62, *p* = 0.212. The median regression length was not significantly different at Time 2 than at Time 1. There was no statistically significant main effect in relation to the between-subject variable *grou*p *F* (1, 0) = .24, *p* = 0.625. In other words, the median regression length, overall, was not statistically significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 0) = .27, *p* = 0.605. In other words, while there was a decrease in the median regression length between Time 1 and 2 for the experimental group, that increase was not statistically significantly different from the decrease in median regression length of the control group.

Total number of regressions

There was a significant main effect in relation to the within-subject variable *time F* (1, 6) = .42, *p* = 0.017. In other words, the total number of regressions was significantly lower at Time 2 than at Time 1. There was no statistically significant main effect in relation to the between-subject variable *group F* (1, 0) = .03) *p* = 0.869. In other words, total number of regressions, overall, was not statistically significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 0) = .18), *p* = 0.675. While there was a decrease in the number of regressions between Time 1 and 2 for the experimental group, that decrease was not significantly greater than the decrease in the number of regressions of the control group.

Proportion of saccades that are regressions

There was a significant main effect in relation to the within-subject variable *time F* (1, 4 =.25, *p* = 0.048. In other words, there was a significant decrease in the proportion of saccades that are regressions from Time 1 to Time 2. There was no statistically significant main effect in relation to the between-subject variable *group F* (1, 0) = .78, *p* = 0.385. In other words, the number of saccades that are regressions overall, was not statistically significantly different across groups. There was no statistically significant interaction effect between *group* and *time F* (1, 1) = .73, *p* = 0.189. In other words, while there was decrease in the proportion of saccades that are regressions between Time 1 and 2 for the experimental group, that increase was not statistically significantly different from the decrease in proportion of saccades that are regressions of the control group. *Figures 5.10* and 5.*11* below show the proportion of saccades that are regressions for the experimental group and control group at Time 1 and Time 2.

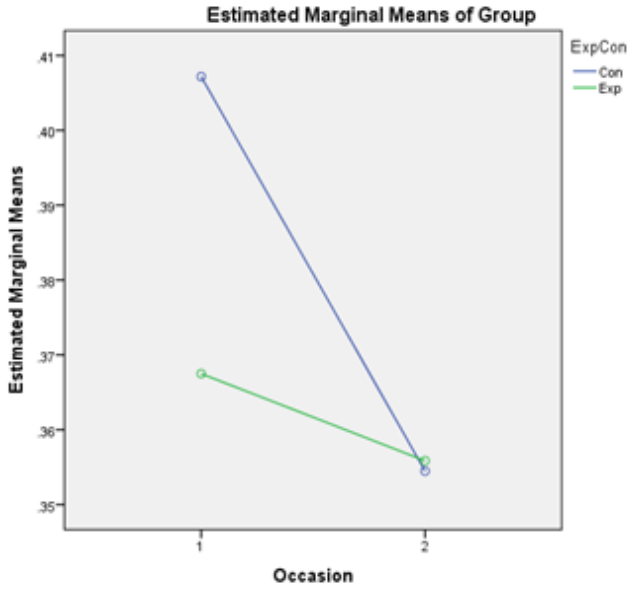


Figure 5.10 Proportion of saccades that are regressions plot

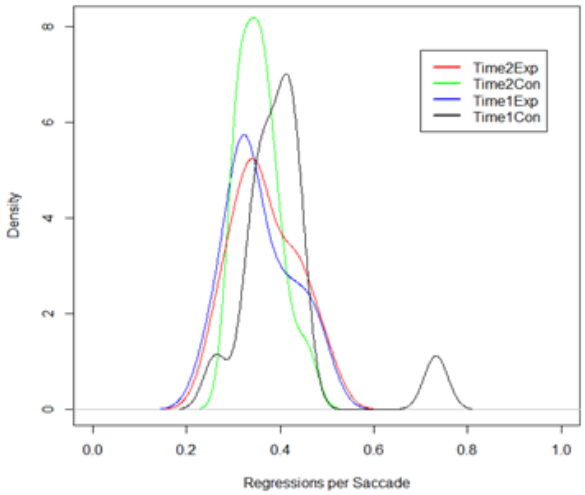


Figure 5.11 Proportions of saccades that are regressions density plot

Vowel analysis

Table 5.9 below presents the sum of visits on vowels and on consonants per person in the experimental and control groups at Time 1 and Time 2.

Table 5.9 *Sum of all visits*

Sum of all visits	**T1**	**T2**
Experimental Consonants	834	686
Experimental Vowels	618	579
Control Consonants	604	523
Control Vowels	530	473

Table 5.10 below presents a summary of the mean number of visits on vowels and on consonants per person in the experimental and control groups at Time 1 and Time 2.

Table 5.10 *Mean of visits*

Mean of visits	**T1**	**T2**	**SD**
Experimental Consonants	46.4	38.1	5.868
Experimental Vowels	34.3	32.2	1.484
Control Consonants	43.2	37.4	4.101
Control Vowels	37.9	33.8	2.899

Analysis of proportion of time spent looking at vowels

There was a significant main effect in relation to the within-subject variable *time* F (1, 6) = .00), p = 0.020. In other words, the participants spent a significantly longer proportion of time looking at vowels at Time 2 than at Time 1. There was no statistically significant main effect in relation to the between-subject variable *group* F (1, 7) = .43, p = 0.011. The proportion of time spent looking at vowels was not statistically significant different across groups. There was no statistically significant interaction effect between *group* and *time* F (1, 3) = .56, p = 0.069. In other words, while there was an increase in time spent looking at vowels between Time 1 and 2 for the experimental group, that increase was not statistically significantly greater than the increase in time spent looking at vowels for the control group. *Table 5.7* below shows the proportion of time spent looking at vowels for the control and experimental groups.

Table 5.11 *Proportion of time spent looking at vowels*

	Time 1	**Time 2**	**SD**
Control	0.47	0.48	0.007
Experimental	0.42	0.45	0.021

5. 7 Discussion

Textual enhancement

The results of Study Two show that textual enhancement did not significantly increase the reading comprehension gain scores of the experimental groups in the pre- and post-reading tests. However, generally speaking, there is little consensus in the findings of enhancement studies. Jahan and Kormos (2014) proposed that textually enhanced forms may not always be noticed by learners because salience created by teachers may not correspond with learners' internally generated salience. Secondly, prior knowledge of the construction being studied may facilitate noticing of the textually enhanced conditions. Additionally, short-term treatment with limited exposure may not elicit the desired effect. It is also possible that statistically significant gains in reading proficiency may require interventions of considerable greater intensity than that provided in Study Two. Finally, it is possible that the lack of significant impact of the intervention may be due to the enhancement itself. In hindsight, highlighting all of the vowels may have been overwhelming for the participants. Enhancing only one vowel letter per reading might have been more effective and might have helped the students to distinguish between the vowels, which the literature suggests may be one of the problems for this population.

Lee (2007) suggested that researchers rely on the participants' institutional status to classify proficiency level and that this could potentially be a factor affecting the ability to focus on

forms. This is another plausible explanation for the results of Study Two where students had to focus on vowels, as in all the classes in the experiment, there was a mixture of students who were placed into the level by a placement test, and those students who had been at the college for several semesters and were placed into the level by successful completion of the previous level. Therefore, students might have had different levels of reading ability which might have acted as an uncontrolled moderator variable in this study.

The question arises as to why the results of my study are in contrast to those of Alsadoon and Heift (2015). In terms of the effectiveness of textual enhancement on intake of vowels as measured by a word recognition task, Al Sadoon and Heift (op.cit.) report that the experimental group outperformed the control group on word form errors and there was a significant gain in word form "implying that the treatment significantly improved the learners' orthographic vowel knowledge and thus reduced their vowel blindness" (p. 69). As I pointed out in Chapter 4, the term "vowel blindness" is inaccurate. In fact, Study One showed that the Arabic L1 EFL students allocated **more** attention to vowels than did the skilled English L1 participants.

Phonological awareness training (PA)

Lesaux and Siegel (2003) concluded that PA is "as effective for ESL children as for English L1 children" (p. 106). Yeung, Siegel and Chan (2013) found that PA instruction facilitated the acquisition of phonological awareness, expressive vocabulary, word reading and also word spelling. However, both these studies were conducted with young children and it has been shown that PA develops only up to a certain age. According to the Simple View of Reading (Gough & Tunmer,1986; Hoover & Gough, 1986; Hoover &Gough,1990) word decoding and word development cease to make a contribution to reading comprehension after a certain age. Therefore,

it is entirely possible that phonological awareness training is less effective in older students. However, anecdotally, I found that students in the experimental groups became very adept at sounding out the words while reading. A pre-and post-test assessing phonological awareness, phoneme-grapheme correspondences, and phonetic decoding skills might have revealed differences between the experimental and control groups from Time 1 to Time 2. Additionally, a PA awareness test at the beginning of the study might have indicated aspects of participants' reading aloud ability that may not have been systematically targeted during the classroom reading instruction with the experimental groups.

Word recognition and automaticity

My results do not support Akamatsu's (2008) findings that learners benefit from word recognition training in speed and accuracy. However, his study was conducted with Japanese university students and this may prevent the findings being generalizable to Arabic L1 college students. As mentioned in Chapter 3, Fender (2003) found that Arabic L1 EFL students were significantly slower than a group of proficiency-matched Japanese students in a lexical decision task. They also experienced more difficulty than other EFL populations in processing English word forms and had slower and less effective context-free word recognition skills.

Segalowitz and Segalowitz (1993) proposed that in the initial stages of L2 word recognition, there is merely a speedup of performance which is qualitatively different from automatization. Fukkink, Hulstijn and Simis (2005) investigated the automatization of lexical access and interpreted their findings as lending support more for an acceleration interpretation than for a qualitative change interpretation of automatization. They did not find a significant improvement in reading comprehension after training. However, reading fluency is achieved

through the development of automaticity and requires extended periods of implicit learning. A large recognition vocabulary size is also necessary for fluent reading performance. This skill set is only learned gradually and is not always easy to detect in shorter training studies (Grabe 2010). The duration of the interventions was possibly not long enough to detect any gains in automaticity. Additionally, automaticity was not directly assessed by the pre-and post-tests. Finally, it is possible that some of the control group's activities i.e. word searches, NGSL list with meanings, and the PHRasal Expressions List (which provided multiple exposures to the vocabulary) may have had a facilitating effect which is not exponentially different from the activities of the experimental group.

Oral text reading fluency

The National Reading Panel (2000) concluded that guided oral reading produced significant and positive impact on word recognition. Jiang et al. (2012) investigated the relationship between oral reading fluency and reading comprehension with Chinese L1 EFL students and found that oral passage reading correlated significantly with comprehension. Rasinski, Rikli and Johnston (2009) reported that oral reading fluency continued to have importance beyond primary and middle grades. Binder et al. (2008) observed that adults with low literacy skills paused longer than skilled adults, had more irrelevant pauses within sentences and stumbled on more words. Surprisingly, my results do not support the findings that oral fluency training contributes to reading comprehension. The brevity of the intervention may be one reason for the findings. Additionally, there were no pre- or post- tests conducted on oral reading fluency (see for example those used by Klauda and Guthrie, 2008) which might have detected changes in passage expressiveness, phrasing, pace, smoothness and word expressiveness over the course of

the intervention. If significant improvement was made in these areas, a case could then be made for the possibility of gains in reading comprehension with a longer intervention.

Spelling intervention

Ehri (1997) and Perfetti (1997) observed that spelling and reading appear to be "both sides of the same coin" even though spelling is more challenging for most students. Incomplete or inaccurate spelling representations will result in less efficient word recognition skills (Ehri, 1997; Perfetti, 1992). Findings have shown that children who are good readers are usually good spellers (Ehri, 1987). Ehri and Wilce (1979) found that students remembered the spoken words better when they had seen the spelling. Ehri and Rosenthal (2007) reported that the effect of spelling is not limited developmentally to the period of beginning to read. Graham, Harris and Chorzempa (2002) examined the effect of supplemental spelling instruction on ESL students and found that the students in the spelling condition made greater improvements in norm-referenced spelling measures, a writing fluency test and a reading word attack measure. The findings of Study Two did not demonstrate a significant contribution of spelling training to reading comprehension. However, there are several possible reasons for this surprising result. As there were no pre-and post-tests of spelling in any of the groups, it is not possible to ascertain if any significant progress was made in spelling by the experimental groups. Additionally, the brevity of the intervention and the low number of words which were included (the first 2,000 of the NGSL) may have been factors in the results.

Test of reading

It is plausible that another instrument designed specifically to measure reading comprehension at the word level, would have yielded different results. The instrument that I used was not capable of detecting any small differences in the reading ability of the participants. Additionally, for a reading test to be valid in an academic setting, it should assess the range and level of cognitive processing at both higher and lower levels. If it tests only a limited range of processes or only low-level cognitive processes, then it cannot claim to be an appropriate tool for assessing the academic language competence required at university level (Bax, 2012; Bax & Weir, 2012). If the items in the instrument had been influenced by the Khalifa and Weir (2009) model of reading with items assessing global versus local and careful versus expeditious reading, the results might have been different. Another shortcoming is the length of time during which the interventions took place. It may have been too short to allow for any significant gains in scores. It was hoped that the pedagogical interventions would initiate a sequence of cognitive processes initiated by noticing. However, the amount of "time needed for them to be set in motion and completed has yet to be empirically ascertained" (Han, Park& Comb, 2008).

Eye tracking measures

The results for the eye tracking measures showed no statistically significant interaction effect for the mean number of fixations, median fixation duration, median saccade length, number of forward saccades, median regression length, and total number of regressions or proportion of saccades that are regressions. The results of the vowel and consonant analysis showed no statistically significant interaction effect between *group* and *time* for the sum of all visits on vowels, sum of all visits on consonants or the proportion of time spent looking at vowels. However,

in the analysis of the proportion of time spent looking at vowels, the p-value approached significance at the 0.05 level. Sample size affects whether a difference between samples is deemed significant or not. In small samples, large differences can be non-significant (Field, 2013). It is plausible that with a larger sample size, the results would have indeed been statistically significant.

The question arises as to why the results of my study are in contrast to those of Alsadoon and Heift (2015) which showed that vowel blindness was significantly reduced for the experimental group due to a longer focus on target words as suggested by their eye tracking data. As I mentioned in Chapter 3, there are several limitations with the data analysis and interpretation of their findings. Their AOIs were drawn around words and not on vowels and consonants. Consequently, there is no evidence that textual enhancement drew the visual attention of the Arabic L1 EFL students' attention to a target word **and** its vowels. Neither is there evidence that participants made longer fixations on a target word **and** its vowels.

5.8 Conclusion

Chapter 5 described Study Two which was an experiment comparing two groups of Arabic L1 EFL students before and after focused reading interventions in the classroom. It assessed changes in reading behaviour and performance by analysing eye movements and using pre-and post-reading tests. Specifically, this part of the study addressed the following research questions:

1) Do focused intervention activities change the eye movement patterns of Arabic L1 EFL students while reading at the word and sentence level in English? If so, how and to what extent?

2) Do focused intervention activities influence word and sentence level reading processes and thus overall reading comprehension? If so, to what extent? With regards to RQ 1, the data showed that focused interventions did not produce statistically significant changes in the eye-movement patterns of the participants in this study. However, these results should be interpreted recognizing

the limitations inherent in the research design of the study as discussed above. The findings of the second RQ revealed that the focused interventions did not produce statistically significant gain scores in reading proficiency as assessed by pre-and post-tests. These findings must also be considered recognizing the limitations of the instrument used to asses reading proficiency.

Chapter 6: Conclusion

Introduction

In this chapter, I will review and summarize the main findings of Study One and Study Two. I will then discuss the theoretical and methodological contributions of the studies and the pedagogical implications. I will then consider the limitations of my research and make suggestions for further studies.

6.1 Main findings

6.1.1 Study One

Study One was an eye tracking experiment which compared the eye movements of skilled English L1 speakers and Arabic L1 EFL students. It was designed to answer the RQ: Do the eye movements of Arabic L1 students at a technological college in Qatar differ from those of skilled English L1 readers while reading English sentences? If so, how do they differ and to what extent? Table 6.1 below shows the metrics calculated and confirmation or negation of the expectations that were outlined in Table 4.1.

Table 6.1 *Metrics calculated and expectations*

Metrics calculated	**Expectations**
Number of fixations	**Confirmed.** The Arabic L1 EFL participants made significantly more (p= <0.00) fixations than the English L1 participants.
Median fixation duration	**Confirmed**. The Arabic L1 group made significantly longer (p= <0.00) fixations than the English L1 group.
Forward saccade length	**Confirmed**. The Arabic L1 group made significantly shorter (p= <0.001) forward saccades than the English L1 group.
Number of forward saccades per individual	**Confirmed**. The Arabic L1 participants made significantly more (p=<0.001) forward saccades than the English L1 participants.
Regression length	**Confirmed**. The Arabic L1 group made significantly shorter (p=<0.001) regressions than the English L1 group.
Proportion of saccades that are regressions	**Not confirmed.** The proportion of regressive movements of the Arabic L1 group was not significantly different from the proportion of regressions of the English L1 group.
Sum of all visits on consonants Mean of all visits on consonants	**Confirmed**. The Arabic L1 group made twice the number of visits on consonants than did the English L1 group.
Sum of all visits on vowels Mean of all visits on vowels	**Not confirmed**. It was the Arabic L1 group which made twice the number of visits on vowels than did the English L1 group.
Proportion of vowel visits	**Not confirmed**. Although there was a significant difference (p=<0.001) between the two groups in the mean proportion of vowels, it was the Arabic L1 group which executed the higher proportion of visits on vowels.

The eye movements of the Arabic L1 EFL students were significantly different from those of the skilled English L1 participants in 8 of the 9 metrics calculated. As was expected, the Arabic L1 EFL participants made significantly more

fixations than the English L1 participants. The number of fixations overall is thought to be negatively correlated with search efficiency. A high number of fixations would be "indicative of difficulty in interpreting the information" (Holmqvist et al, 2011, p. 413). The Arabic L1 participants also made significantly longer fixations than the English L1 participants. Generally speaking, beginning readers, poor readers and dyslexic readers have longer fixations (Rayner, 1998; Ashby, Rayner & Clifton, 2005). With regards to forward saccades, the Arabic L1 EFL participants made significantly shorter saccades and consequently executed significantly more saccades than the English L1 participants. As a rule, as ability increases, the number of forward saccades decreases, while the length of the saccades increases (Rayner, op.cit). The Arabic L1 EFL participants also made significantly shorter regressions than the English L1 participants. As ability increases, the length of regressions increases (Rayner, op.cit). The proportion of saccades that are regressions was not significantly different between the groups. However, as mentioned in Chapter 4, this result may be an artefact of the task itself, as research has shown that readers make more regressions to a previously read passage for re-inspection when this segment helps answer a subsequent question (Christie & Just, 1976; Kennedy & Murray, 1979; Kennedy et al., 2003). It is therefore possible that the English L1 participants may have been encouraged by the task to make more regressions than they would normally have made.

Viewed collectively, there are robust differences between the eye movements of the Arabic L1 and the English L1 participants.

Study One provides a preliminary explanation for the specific difficulties experienced by Arabic L1 students learning to read in English. It would seem that the eye movement patterns displayed by these learners are potentially indicative of less efficient cognitive processes than those displayed by the skilled L1 participants. It might be argued that such differences in eye movements may be manifested by readers from other language background groups. However, the results of Study One corroborate the work of previous researchers who have found that Arabic L1 participants showed more difficulties than did participants from other first language backgrounds. For example, Ryan and Meara (1991) found that performance on tasks by native speakers was less accurate and slower than performance by EFL learners with non-Arabic language backgrounds or native English speakers. In particular, the Arabic L1 participants had a higher error rate in judging deleted vowel stimuli than the participants from other language backgrounds. Hayes-Harb (2006) concluded that the Arabic groups' response time to vowel deletion tasks were significantly slower than those of the non-Arabic groups in her study. In a letter detection test, the Arabic group exhibited the least accurate performance of all the different groups. She concluded that because the performance by the Arabic L1 speakers on the letter detection task "differed so markedly from performance by the other groups, the findings cannot simply be attributed to word processing differences between native and non-native readers of English" (p. 335). Fender (2003, 2008) concluded that Arabic ESL students experienced more difficulty than other ESL populations in processing English word forms. He found that the Arabic speakers were significantly

slower than a group of proficiency-matched Japanese speakers in a lexical decision task.

The most surprising result of Study One was the outcome of the vowel and consonant data. There was a statistically significant difference between the two groups in the proportion of visits on vowels. However, it was the inverse of what had been expected, with the Arabic L1 EFL participants making a higher proportion of visits on vowels. These results provide potential preliminary evidence to question the 'vowel blindness' hypothesis. Ryan (1997) proposed that vowel blindness is a condition seen in Arabic L1 speakers "which may be due to a lack of awareness of the function vowels perform in English". (p. 189). However, as Sadhwani (2006) pointed out, "having problems in recognizing vowels in words where they have been deleted does not seem a reasonable explanation for such a condition." (p. 43). In her (2005) study, she found evidence in a word dictation task which showed that the Arabic L1 speakers did in fact use and recognize the significance of vowels in written English. In Study One, if we take measures of fixation duration and number of visits on vowels as indications of increased attentional processing, we can claim that the Arabic participants allocated more attention to vowels than the English L1 participants. It is my considered opinion that the Arabic L1 speakers are not 'blind' to the vowels in English but are perhaps unsure of how to process them because of what Ryan (1997) termed "the excessive amount of information present in English where all the vowels are written down, a convention which is markedly different from Arabic" (p. 186). This is certainly plausible when comparing the 2 diphthongs and 6 vowels in Arabic to the 14-20 (depending upon regional variations) vowel phonemes

in English. One of the issues revealed in Sadhwani's (2005) study was that participants were not always aware of the accurate sounds that the English vowels represented. If we look at the role of phonological awareness to reading, this lack of sound symbol correspondence would certainly contribute to lack of automatic processing, making the decoding of vowels effortful. I therefore suggest discontinuing the use of the term "vowel blindness" and instead, refer to the phenomenon as a perceived "vowel uncertainty".

6.1.2 Study Two

Study Two utilized a barrage of pedagogical interventions to investigate the following Research Questions:

1) Do focused intervention activities change the eye movement patterns of Arabic L1 EFL students while reading at the word and sentence level in English? If so, how and to what extent?

2) Do focused intervention activities influence word and sentence level reading processes and thus overall reading comprehension? If so, to what extent?

With regard to the eye movement patterns at Time 1 and Time 2:

1) The mean number of fixations, overall, was not significantly different across groups. While there was a decrease in mean number of fixations between Time 1 and 2 for the experimental group, this decrease was not significantly greater than the decrease in mean number of fixations of the control group.

2) The median fixation duration, overall, was not statistically significantly different across groups. While there was a decrease in fixation duration between Time 1 and 2 for the experimental group, that decrease was not significantly greater than the decrease in fixation duration of the control group.

3) The median saccade length, overall, was not statistically significantly different across groups. Although there was an increase in median saccade length between Time 1 and Time 2 for the experimental group, that increase was not significantly greater than the increase for the control group.

4) The number of forward saccades was not statistically significantly different across groups. While there was a decrease in the number of forward saccades between Time 1 and 2 for the experimental group, that decrease was not statistically significantly different from the decrease in number of forward saccades of the control group.

5) The median regression length was not significantly statistically different across groups. Although there was a decrease in the median regression length between Time 1 and 2 for the experimental group, that increase was not statistically significantly different from the decrease in median regression length of the control group.

6) The total number of regressions was not statistically significantly different across groups. While there was a decrease in the number of regressions between Time 1 and 2 for the experimental

group, that decrease was not significantly greater than the decrease in the number of regressions of the control group.

7) The proportion of saccades that are regressions was not statistically significantly different across groups. While there was decrease in the proportion of saccades that are regressions between Time 1 and 2 for the experimental group, that increase was not statistically significantly different from the decrease in proportion of saccades that are regressions of the control group.

8) The proportion of time spent looking at vowels was not statistically significant different across groups. While there was an increase in scores between Time 1 and 2 for the experimental group, that increase was not statistically significantly greater than the increase in scores of the control group.

9) The proportion of time spent looking at vowels was not statistically significant different across groups. While there was an increase in time spent looking at vowels between Time 1 and 2 for the experimental group, that increase was not statistically significantly greater than the increase in time spent looking at vowels for the control group.

Table 6.2 *Features and expectations*

Feature	**Expectations**
Total reading score	**Not confirmed**. There was no statistically significant interaction between *group* and *time.*
Main Idea	**Not confirmed**. There was no statistically significant interaction between *group* and *time*.
Guessing meaning from context	**Not confirmed**. There was no statistically significant interaction between *group* and *time*.
Pronoun reference	**Not confirmed**. There was no statistically significant interaction between *group* and *time.*
Reading for specific information	**Not confirmed.** There was no statistically significant interaction between *group* and *time.*

With regards to the reading proficiency scores of participants at Time 1 and Time 2:

1) Total reading scores. Participants in both groups performed significantly better at Time 2 than at Time 1. However, there was no statistically significant interaction between *group* and *time*. Although there was an increase in score between Time 1 and Time 2 for the experimental group, this increase was not significantly greater than the increase in score of the control group.

2) Main idea. Participants in both groups performed significantly better at Time 2 than at Time 1. However, there was no significant interaction between *group* and *time*. While there was an increase in scores between Time 1 and Time 2 for the experimental group, this increase was not significantly greater than the increase of the control group.

3) Guessing Meaning in Context. Neither group performed significantly better at Time 2 than at Time 1 and there was no statistically significant interaction between *group* and *time*. While there was an increase in score between Time 1 and Time 2 for the experimental group, this increase was not significantly greater than the increase in scorers for the control group.

4) Pronoun reference. Participants in both groups performed significantly better at Time 2 than at Time 1. However, there was no statistically significant interaction between *group* and *time*. While there was an increase in scores between Time 1 and Time 2 for the experimental group, this increase was not significantly great than the increase in scores for the control group.

5) Reading for specific information

There was no significant main effect with relation to the within-subject variable, time $F(1, 1) = .479$, $p = 0.493$. In other words, the participants did not perform significantly better on Time 2 than they did on Time 1. There was no statistically significant main effect in relation to the between-subject variable, group $F(1, 0) = .005$, $p = 0.945$. In other words, the scores, overall, were not statistically significantly different across groups. There was no statistically significant interaction effect between group and time $F(1, 4) = .334$, $p = 0.567$.

These results were unexpected and to some extent disappointing, as it was hoped that the interventions would produce significant changes in the eye movements of the participants in the experimental group and increase their level of reading proficiency. However, the results must be interpreted in light of several limitations to the study which will be discussed in detail in 6.5. However, what is intriguing is the effect of reading task type on the reading scores. As previously mentioned, the participants' scores improved from Time 1 to Time 2 in the Main Idea, Pronoun Reference, and Short Answer tasks, although the increases in scores were not statistically

significant. However, participants did not perform better on the Vocabulary in Context, Multiple Choice, and True/ False /Not Given tasks at Time 2 and, in fact, the control group performed worse on the True/False/Not Given task. I will discuss each of these three tasks in turn.

Guessing Meaning in Context

For many years, reading research has promoted the idea that learners will acquire vocabulary more successfully if they are encouraged to guess meaning from context. Numerous studies have investigated the effect of such *incidental vocabulary acquisition* which Schmitt (2010) defined as learning which occurs as a result of using language with no particular intention to learn a particular linguistic element (see for example, Godfroid et al., 2017; Pellicer-Sánchez, 2015 Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010). These studies confirmed the benefits of seeing target words repeatedly. Pellicer-Sánchez (2016) reported that 8 is the optimum number of times required for a reader to see the target words in order to achieve a more fluent reading behaviour. The Meaning in Context activities used for both classroom and assessment, mandated by the Assessment Department at the college at the time of this research, consisted of students reading a passage once and attempting to guess the meaning of new words from the contextual clues. This would not qualify as a route to incidental vocabulary acquisition. Additionally, it cannot be presumed that all students will be equally effective using this strategy. Pellicer-Sánchez (2013) pointed out that research has shown wide-ranging differences in the ability of learners to guess from context when they are reading.

Ryan (1997) asserted that guessing meaning from context may be precisely the least appropriate strategy for Arabic-speaking learners of English. The task assumes the readers are able to decode accurately most of the words on a page but "if readers have a serious deficit in the ability

to decode items at the word level, top-down processing will break down, and they will be wholly dependent on bottom-ruled processing" (p. 187). Laufer and Sim (1985b) suggested that vocabulary size was a main variable which might negatively affect guessing from context because learners with low sizes of vocabulary are unable to utilize it effectively. The Guessing in Meaning in Context task is potentially an ineffective strategy to teach Arabic L1 EFL students whose vocabulary size is small and it is perhaps not an appropriate test of their reading ability. The nature of the task itself may be the explanation for the low scores at Time 2 as students would not be expected to improve significantly.

Multiple Choice

The processing options available to the reader as presented by Urquhart and Weir (1998) are global versus local and careful versus expeditious reading. The Multiple Choice task in Study Two required the participants to read at the global as well as the local level and therefore participants needed to incorporate information from different parts of the text. They were also required to read carefully, meaning reading the text to obtain complete meaning at either the global or local level.

In retrospect, this particular task is not appropriate to assess reading proficiency at the word and sentence level as per my research question and items should have been revised before they were used in Study Two.

True/False/Not Given

In the True/False/Not Given task, the control group performed worse at Time 2 than at Time 1. Students at the College rely heavily on guessing strategies in assessments to compensate for their reading difficulties. The experimental group had been taught during the semester to refrain

from guessing and to read every word in the question. The control group received no such instruction. It is possible that they continued with their random guessing strategies and remained even more unsuccessful as they had been at Time 1.

6.2 Theoretical and methodological contribution of the studies

The main aim of Study One was to investigate the differences between the eye movements of Arabic L1 speakers and skilled English L1 speakers while reading in English at the word and sentence level. My results showed that there were robust differences between the two groups in terms of number of fixations, fixation duration, forward saccade length, number of forward saccades per individual and length of regressions. The findings help to explain the difficulties that these particular students have when learning to read in English as documented by, for example, Abu-Rabia, 1997b; Fender, 2003, 2008; Hayes-Harb, 2006; Randall and Meara, 1988; Randall, 2007; Randall & Groom, 2009; Ryan, 1997; Ryan and Meara, 1991; Saigh & Schmitt, 2012; Thompson-Panos and Thomas-Ruzic, 1983). The eye movements of the Arabic L1 students in Study One are similar to beginning and poor English L1 readers as well as to dyslexic readers. These findings constitute an important contribution to the literature on the reading difficulties of Arabic L1 students and reveal how 'far away' the eye movements of the Arabic L1 participants are from those of skilled English L1 readers.

Study One is the first study to use eye tracking technology to investigate attentional focus on vowels and consonants with the AOIs drawn on individual letters. This level of specification allowed me to calculate precisely how long participants spent attending to each letter in the words. When the AOIs are drawn on words and not on individual letters, it is not possible to obtain exact measurements. The information provided by the analysis of time spent on vowels versus time spent

on consonants is of particular importance to our understanding of how Arabic L1 speakers process words. The finding that the Arabic L1 participants actually spend more time on vowels than do the English L1 participants is surprising. It provides preliminary evidence to question the hypothesis of 'vowel blindness'.

Bax (2013) pointed out that much of the eye tracking research has been conducted with L1 participants and results may therefore not be relevant to L2 reading. Nassaji (2012) suggested that results from the vast amount of SLA research cannot be used by classroom teachers because the research is generally carried out under controlled experimental conditions and results cannot therefore be generalizable. Additionally, many researchers in SLA may not be familiar with classroom issues and may not investigate areas which teachers consider to be meaningful. Study Two addresses these three concerns. First, it was carried out with L2 readers and as such was not in the 'default mode' of reading (Reichle, Warren & McConnell, 2009), where comprehension proceeds without difficulty. Secondly, the study was carried out in real classrooms with the regular classroom teachers over a standard semester with students typical of other Gulf countries, making the results generalizable to other classrooms in this area. Finally, the impetus for Study Two stemmed from a very practical classroom issue: the difficulties my students have learning to read effectively in English. This is a very real problem faced by all teachers in the Middle East who teach reading to Arabic L1 students. To my knowledge, Study Two is the first empirical study to investigate the effects of a barrage of pedagogical interventions on the eye movements and reading proficiency of Arabic L1 EFL students. The interventions were integrated easily into the classrooms of the experimental groups and included both input (enhanced texts, tracking exercises,) and output (rapid word naming, oral fluency, spelling and phonological awareness)

tasks. Although the interventions did not produce statistically significant results, the study provides a building block for future studies using pedagogical interventions with these particular learners.

6.3 Pedagogical implications

Diagnosis is intended to lead to treatment, and the more specific the diagnosis is, the more likely it is that effective remediation can be developed (Harding, Alderson & Brunfaut, 2015). The implication is that teachers who can identify their students' needs can adapt their instruction to meet those needs. Study One generated very useful information as to how vastly different the Arabic L1 EFL participants' eye movements are from those of skilled English L1 readers. In fact, the results showed that the Arabic L1 EFL students' eye movements are similar to those of beginning, poor, and dyslexic readers. The current practices employed in the majority of reading classes in the Gulf region do not seem to be successfully preparing the students to read in an effective and efficient manner. By taking best practices from the literature on reading difficulties, the interventions for Study Two were chosen. Without word decoding and word recognition strategies, effective reading is not possible.

Study Two looked at the effectiveness of these pedagogical interventions implemented specifically to address the problems that Arabic L1 EFL students encounter when learning to read in English.

Fluency

According to Rasinski (2010) reading fluency refers to the ability to "read the words in a text effortlessly and efficiently (automaticity) with meaningful expression that enhances the meaning of the text cognition (prosody)" (p. 31). It is evident that the Arabic L1 EFL students are

far from being fluent readers and this would indicate that there needs to be a shift in the way we teach these students how to read. Unfortunately, fluency is widely equated with reading rate and in many classrooms, there is an emphasis on speed as the reading goal. As previously mentioned, in typical classrooms in the Gulf "little actual reading occurs, with most of the time devoted to tasks and activities that assume the reading of the text" (Grabe, 2009, p. 379). Most commonly, teachers assign a text, students read it silently and comprehension is assessed by answering the questions which follow. Pikulski and Chard (2005) described fluency as a bridge from word recognition accuracy to text comprehension. It is not enough for readers to read a text accurately, they need to read the words automatically and they develop automaticity through wide and deep practice (Rasinski, 2012). It is crucial that teachers of Arabic L1 EFL learners incorporate 'deep' or what is often referred to as repeated reading (Samuels 1979) in their classrooms. Deep reading is a classroom practice during which a learner is asked to read a single text repeatedly until a level of fluency is achieved. Research showed that word recognition accuracy, automaticity, comprehension and attitude toward reading improved with deep or repeatedly reading (see, for example Kuhn & Stahl, 2003; Rasinski, et al., 2011.)

"If automaticity is the fluency link to word recognition, prosody completes the bridge by linking fluency to comprehension" (Rasinski, 2012, p. 51). Prosody and automaticity must work in concert and just as automaticity develops through wide and deep reading, so too does prosody. However, despite the emphasis on reading speed in some classrooms, the goal of wide and deep reading is not to increase speed but to achieve an expressive oral reading of the passage that enhances the text (Rasinski, op.cit.). The National Reading Panel (2000) undertook a comprehensive review of the major components found to contribute to skilled reading. They focused on research pertaining to word recognition, (phonological awareness, learning the

alphabetic principle, vocabulary, fluency and comprehension). In addition to word recognition, they reviewed research on fluency and comprehension. One approach that they found to be highly effective was guided repeated oral reading which encourages students to read passages orally with systematic and explicit guidance and feedback from teachers. In the past, interventions focused on reading speed as the main goal. However more recent research has focused on the role that prosody plays in reading and how listening to an audio model of the text may act as a scaffold to reading comprehension (Brevik, Olsen & Hellekjaer, 2016). If the learners are consciously aware of their word identification difficulties, or if they can be taught to be aware of them, it is possible that these conscious strategies may help them process English words (Hayes-Harb, 2006, p. 337). It is imperative that all reading classes with Arabic L1 EFL students embed automaticity and fluency activities into their everyday curriculum. This leads to the next pedagogical implication of this study.

Teacher education and pedagogical development

High quality reading instruction is partially defined by the knowledge that teachers of reading must possess to provide effective instruction for their students (National Reading Panel, 2000). Research suggests that teachers of reading, like teachers of physics or chemistry, require domain-specific knowledge and expertise. This is often referred to as pedagogical content knowledge (Shulman, 1987). However, many teachers do not possess the knowledge and skills needed to provide quality reading instruction. The assumption is that being a skilled reader creates a sufficient knowledge base for providing reading instruction. As compared with disciplines such as mathematics and social studies, studies of teachers' pedagogical content knowledge are not well developed (Cunningham & Ryan O'Donnell, 2015).

Mather, Bos and Babur (2001) investigated the impact of years of experience on teaching knowledge and attitudes. They found that pre-service and in-service teachers did not have a clear understanding of the importance of letter-sound correspondences, believing that the use of context was the most beneficial strategy for identifying an unknown word. Moats (1994) found that neither pre-service nor in-service teachers had sufficient knowledge of the phonological, orthographic and morphological structures of English to effectively teach reading at a basic code level. She was one of the first to suggest that many teachers "understood too little about spoken and written language structure to be able to provide sufficient instruction in these areas (p. 81). Some studies (see for example Cunningham, Perry, Stanovitch & Stanovitch, 2004; Cunningham & Zibulsky (2009)) showed that teachers have difficulty counting the number of phonemes and morphemes in words and in classifying words by syllable type (open, closed, r-controlled, silent e, vowel combination, and consonant-le). The National Council on Teacher Quality (NCTQ) (2006) examined what pre-service teachers learn about reading instruction in their teacher preparation programme and concluded that teacher education programmes are not consistently teaching the principles and practices that recent evidence has demonstrated to be effective. They reported that not only are the majority of teacher candidates not receiving sufficient exposure to scientifically based methods of reading instruction but that teacher candidates are often advised to develop their own unique approach to the teaching of reading (Cunningham & Ryan O'Donnell, 2015). The teaching of reading in a second or foreign language fares even worse. Long (2014) decries the situation of Language Teaching (LT).

> "LT, unfortunately lacks the characteristics of a true profession, such as law, medicine, engineering, nursing or architecture…..Among institutions offering courses or whole degree programs supposedly preparing students for a career in LT, there is no agreed-upon

common body of knowledge of which all practitioners should demonstrate mastery and no common examinations required of would-be practitioners. There is no licensing body, no licenses and few sanctions on cowboy teachers or language schools" (p. 6).

The NRP (2000) argued that the ability to create the ideal combination of teaching techniques requires a deep understanding of reading development and a familiarity with the pedagogical strategies which have been proven to be most effective in supporting reading development. This would necessarily include also the understanding of the facilitative effects of repeated reading exposure across a wide variety of texts. They stated that reading teachers must be able to provide quality instruction in five areas: phonological awareness, phonics, fluency, comprehension, and oral language development including vocabulary.

If we acknowledge that many teachers lack the skills to provide quality reading instruction, we must then consider how to address these gaps. The Status Report on Teacher Development in the United States and Abroad (2009) concluded that in education, professional development is flawed because teachers lack time and opportunities to observe each other's classrooms, learn from mentors and work collaboratively. Some of the key findings in their report which are pertinent to this study are that:

1) sustained and intensive professional development for teachers is related to students' achievement goals.

2) effective professional development is intensive, ongoing and focuses on the teaching and learning of specific academic content.

3) teachers typically need substantial professional development in a given area (50 hours) to improve their skills.

It sometimes occurs that when language teachers receive feedback which is intended to increase their efficacy, they construe it as a threat and choose consciously to resist change or innovation (Hiver & Dörnyei, 2015). "Ultimately this well-documented aversion to change may interfere significantly with the language teacher's ability to meaningfully reflect on their practice and develop as a professional, with potentially detrimental consequences for learners they come in contact with" (Hiver & Dörnyei, op. cit. p. 15).

I would recommend that schools, colleges and universities in the State of Qatar invest more time and financial resources to ensure that teachers have sufficient exposure to evidence-based methods of teaching reading. I also recommend that any pedagogical development implemented for in-service teachers be elective. If participation is optional, teachers will not resist innovation and will be more likely to create environments conducive to language development" (Hiver & Dörnyei, 2015).

Texts

As mentioned in Chapter 2, the Arabic L1 learner of English does not appear to experience optimum success with mainstream reading materials, despite the claims by writers and publishers that these textbooks can be used in any geographical location with learners of various language backgrounds. As Grabe (2009) stated, a major concern with the adoption of mainstream textbooks is that they are often presented as the "answer" for teaching reading rather than as a resource. Oftentimes, in the Gulf region, the text is the curriculum. Unfortunately, textbooks which focus mainly on the relatively higher-level skills of skimming and scanning do not address the particular difficulties experienced by the Arabic L1 EFL student. As Ryan (1997) stated, it is appropriate to teach reading skills at an advanced level through skimming or scanning if the teacher has confidence in the learners' lower-level skills to allow them to take risks. Such strategies such as

guessing meaning from context work well with learners who share a considerable amount of cognate vocabulary with English. However, it appears that it may be "more difficult for Arabic L1 speakers since it is precisely at the word level that their difficulties arise and context does not always help much" (p. 188). As mentioned in Chapter 5.3, the textbooks used in the college where the studies took place focus on relatively higher-level skills such as skimming. There are no automatic word recognition activities and no fluency exercises found in any of the levels. Equally important, there are very few reading passages for students and teachers to exploit. For example, in the Unlock Series (Cambridge University Press), there are only two reading passages per unit, over 10 Units. In order to assist the Arabic L1 EFL learner in the Gulf, a concerted effort needs to take place whereby qualified personnel create in-house materials to provide our students with the kinds of reading activities and support they so desperately need. We cannot continue to proceed with reading instruction in a 'one size fits all' approach.

6.4 Limitations

Studies One and Two are not without limitations and, as previously mentioned, the results must be interpreted in light of these.

Study One

Firstly, there is a sampling limitation. Both the Arabic L1 and English L1 participants comprised a convenience sample elicited from the faculty and students at the College. Additionally, there were only 75 participants in total; 36 English L1 and 39 Arabic L1. In future studies, if logistically conceivable, access to a larger number of participants randomly selected would provide more robust data.

Secondly, there could potentially be a limitation regarding the actual task that the participants were asked to carry out after looking at the signs. They were asked to decide which of the following two sentences correctly described the sign. In the analysis, there was no statistically significant difference between the two groups in terms of the proportion of saccades that were regressions. However, as mentioned in Chapter 4, previous research reported that readers tend to make more regressions to a previously read passage for re-inspection when this segment helps answer a subsequent question (Christie & Just, 1976; Kennedy & Murray, 1987). Therefore, the English L1 speaking participants might have been encouraged to make more regressions than they would normally make. An alternative task may have produced different results. However, the trade-off would be that participants would not have had a reason to read the signs that were presented to them.

Study Two

Study Two has limitations that are inherent in any Exploratory Practice (EP) study. Firstly, there was a sampling limitation, both in terms of participant numbers and in terms of the participants being a convenience sample. However, in classroom research such as this, it is not possible to have access to large numbers of participants who are randomly chosen. Another limitation which is often present in Exploratory Research is the length of time that participants are involved in the study. When conducting research in an educational institution, the researcher has no control over the length of time students spend in class or the length of terms and semesters. Thus, a further limitation of Study Two is the length of time over which the pedagogical

interventions could be implemented as progress is not always reflected in studies of a shorter duration. Lastly, again because of educational restraints, it was not possible to conduct delayed post-tests which may have given insight into any long-lasting learning gains. However, conducting research in educational settings under regular conditions makes the results more generalizable than a study carried out in an ideal laboratory setting.

No pre-treatment or post-treatment assessments were administered to assess rapid word recognition, phonological awareness, spelling or oral text reading. Such evaluations would have given an insight into the gains that were made in the individual skill sets that comprise the reading process. Considerable improvement may have taken place in some or all of the skills but this could not be not demonstrated.

The last and most serious limitation to Study Two is the mismatch between Research Question 2 and the instrument used to assess it. My research question asked whether focused intervention activities influence word and sentence processes and thus overall reading comprehension. However, the pre-and post-test used to determine this had no tasks that could demonstrate if and how word and sentence level processes were influenced.

6.5 Future research directions

I envisage three potential future research directions. The first one would be to build on the preliminary findings of Study One. As mentioned previously, all the participants in Studies One and Two were in the same EFL level at the technological college and were all between the ages of 18 and 26. Therefore, one potentially informative research project would be to investigate the eye movements of students at different CEFR levels to ascertain if and to what extent the eye movements of participants approximate those of skilled English L1 readers as proficiency

increases. Additionally, it would be of interest to investigate the eye movements of different age groups of Arabic L1 EFL students to determine if and to what extent their eye movements change as age increases.

To build on the findings of Study Two and to address the limitations of this study, another future project would be to investigate the effects of focused pedagogical interventions specifically on word recognition (phonological awareness, spelling, automaticity) and on fluency. Pre-and post-tests of the components of word recognition would provide potentially useful information about their growth throughout the intervention period. At a later stage, the effects of these interventions in terms of reading proficiency could be measured.

A third project would be to determine if the EFL teachers in the school system in Qatar have the requisite knowledge of the structure of spoken and written English. In Moats' (1994) study, after they were given a survey of their knowledge of phonemes and morphemes and how these elements are represented in writing, the self-selected participants took a course which focussed on phonemic awareness training, spoken-written language relationships, and careful analysis of spelling and reading behaviour in children. After the course, the teachers judged the course content to be essential for teaching reading and suggested that it become a prerequisite for certification. The self-selection aspect eases the resistance to change and innovation that often accompanies compulsory pedagogical development sessions. Moats (op.cit) noted:

> "Until we recognize that teachers do not naturally acquire the kind of expertise in language structure that is required of them for remediating and preventing reading problems, we will neglect to provide the necessary training. Teachers will continue to teach without understanding the alphabetic orthography and how it represents speech, will continue to

give students misinformation, and will be unable to clarify concepts or to organize language instruction beneficially. Lower-level language mastery is as essential for the literacy teacher as anatomy is for the physician. It is our obligation to enable teachers to acquire it" (p. 99).

I believe that all our Arabic L1 EFL students in Qatar deserve to learn to read effectively and efficiently before they arrive at our technological colleges and universities.

6.7 Concluding remarks

Study One provided preliminary results indicating that the students' eye movements were significantly different from those of skilled English L1 readers. Study One also presented findings that question the 'vowel blindness' hypothesis (Ryan & Meara 1991). The data showed that the Arabic L1 EFL participants spent more time attending to the vowels then did the English L1 participants. I have therefore proposed that the Arabic L1 students are not 'blind' to, but more likely uncertain as to how to process all the information that the vowels carry in English.

Study Two investigated the effects of a barrage of focused pedagogical interventions on the eye movements and reading proficiency of Arabic L1 EFL students. The findings revealed no statistically significant difference between the experimental and control group. However, this preliminary study provides the building blocks for further studies implementing focussed reading interventions in the EFL classroom her in Qatar.

Motivation for this research stemmed from a practical classroom issue: how to remediate the difficulties my Arabic L1 students encountered while learning to read in English.

The aim of the Qatar National Vision 2030 plan is to transform Qatar into an advanced society capable of achieving sustainable development by 2030. The plan's development goals are

divided into four central pillars: human, social, environmental and economic development (General Secretariat for Development and Planning, 2008).

1) Human development – the development and promotion of education among all to create a sustainable and prosperous society.

2) Social development – the development of a just and caring society, capable of playing a key role in establishing global partnerships.

3) Economic development – the development of a diversified economic base to secure and maintain a high standard of living in the future.

4) Environmental development – to maintain a balance between economic and social development and ways of protecting the environment

If this vision is to be achieved, literacy in English is paramount for all of Qatar's citizens. We can no longer ignore the fact that the methods being used to teach reading in English in the schools are inadequately preparing our students for the challenges that lie ahead. It is my sincere hope that this and future research will play a part in starting change and innovation to ensure that the students in Qatar have access to quality English language education so they may be able to contribute to their country's sustainable future.

References

Abu Rabia, S. (1995). Learning to read in Arabic: Reading, syntactic, orthographic and working memory skills in normally achieving and poor Arabic readers. *Reading Psychology: An International Quarterly, 6,* 351-394.

Abu-Rabia S. (1997). Reading in Arabic orthography: The effects of vowels and context on reading accuracy of poor and skilled native Arabic speakers. *Reading and Writing, 9,* 65-78.

Abu-Rabia, S. & Siegel, L. (2003). Reading skills in three orthographies: The case of trilingual Arabic–Hebrew–English-speaking Arab children. *Reading and Writing: An Interdisciplinary Journal, 16,* 611-634.

Ackerman, R. & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology: Applied, 17*, 18-32.

Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics, 29,*175-193.

Alderson, J. (1978). *A study of the cloze procedure with native and non-native speakers of English* (Unpublished doctoral dissertation). University of Edinburgh, UK.

Alderson, J. (1979). The cloze procedure as a measure of proficiency in English as a foreign language. *TESOL Quarterly, 13,* 219-227.

Alderson, J. (1984). Reading in a foreign language: A reading problem or a language problem? In J. Alderson & A. Urquhart (Eds.), *Reading in a foreign language* (pp. 1-27). London, UK: Longman.

Alderson, C., Bastien, S. & Madrazo, A. M. (1977). *A comparison of reading comprehension in English and Spanish*. Research and Development Unit (Report No. 9). Mexico City: UNAM.

Alderson, J., Brunfaut, T. & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics, 36,* 236-260.

Al Fadala, A. (2015*). K-12 reform in the Gulf Cooperation Council countries: Challenges and policy recommendations*. Symposium conducted at the meeting of The World Innovation Summit for Education (WISE), Qatar Foundation, Doha, Qatar.

Allington, L. (1983). Fluency: The neglected reading goal. *The Reading Teacher, 36,* 556–561.

Allwright, D. (1993). Integrating research and pedagogy: Appropriate criteria and practical possibilities. In J. Edge & K. Richards (Eds.), *Teachers develop teachers' research* (pp. 125-135). Oxford: Heinemann.

Allwright, D. (2003). Exploratory practice: Rethinking practitioner research in language teaching. *Language Teaching Research, 7,* 113-141.

Allwright, D. (2005). Developing principles for practitioner research: The case of exploratory practice. *Modern Language Journal, 89,* 353-366.

Alsadoon, R. & and Heift, T. (2015). Textual input enhancement for vowel blindness: A study with Arabic ESL learners. *The Modern Language Journal, 99,* 57-79.

Anastasiou, D. & Griva, E. (2012). Morphological processing strategies: An intervention for spelling difficulties in English language. *English Language Teaching, 5,* 15-23.

Angele, B., Tran, R. & Rayner, K. (2013). Parafoveal-foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology, 39,* 525-538.

Apel, K. & Masterson, J. (2001). Theory-guided spelling assessment and intervention: A case study. *Language, Speech, and Hearing Services in the Schools, 32,*182 -195.

Ashby, J. (2006). Prosody in skilled silent reading: Evidence from eye movements. *Journal of Research in Reading, 29*(3), 318-333.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58A*, 1065-1086.

Ashby, J., Treiman, R., Kessler, B. & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology, Learning Memory and Cognition, 32,* 416-424.

Baayen, R., Piepenbrock, R. & and Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM), Linguistic Data Consortium. Philadelphia: University of Pennsylvania.

Baccino, T. (2004). *La lecture électronique: De la vision à la compréhension*. Grenoble: Presses Universitaires de Grenoble.

Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A. & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Balota, D., Pollatsek, A. & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*, 364-390.

Barik, H. & Swain, M. (1978). Evaluation of a French immersion program: The Ottawa study through grade five. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 10*, 192-201.

Barker, T., Torgesen, J. &Wagner, R. (1992). The role of orthographic processing skills on five different reading tasks. *Reading Research Quarterly, 27*, 335–345.

Baron, N. (2015). *Words Onscreen: The fate of reading in a digital world*. Oxford: Oxford University Press.

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing, 4*,441-465.

Bax, S. & Weir, C. (2012). Investigating learners' cognitive processes during a computer-based CAE Reading test. *Cambridge ESOL: Research Notes, 47*, 3 -14.

Benjamin, R. & Schwanenflugel, P. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly, 45*, 388-404.

Bernhardt, E. (2000). Second-language reading as a case study of reading scholarship in the 20th century. In M. Kamil, P. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 791-811). New Jersey: Lawrence Erlbaum Associates.

Berninger, V., Vermeulen, K., Abbott, R., McCutchen, D., Cotton, S., Cude, J., Dorn, S. & Sharon, T. (2003). Comparison of three approaches to supplementary reading instruction for low-achieving 2nd grade readers, *Language, Speech and Hearing Services in Schools, 34*(2), 101-116.

Best, R., Floyd, R. & McNamara, D. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*, 137–164.

Bhattacharya, A. & Ehri, L. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities, 37*, 331–348.

Bialystok, E. (2001). *Bilingualism in development, language, literacy and cognition*. New York, NY: Cambridge University Press.

Binder, K., Tighe, E., Jiang, Y., Kaftanski, K., Qi, C. & Ardoin, S. (2013). Reading expressively and understanding thoroughly: An examination of prosody in adults with low literacy skills. *Reading and Writing, 26*, 665-680.

Blignaut, P. & Wium, D. (August 2013). The effect of mapping function on the accuracy of a video-based eye tracker. *Proceedings of the Eye Tracking South Africa Conference (ETSA)*, (pp. 29-31). Cape Town, South Africa.

Block, D. (2000). Revisiting the gap between SLA researchers and language teachers. *Links & Letters, 7*, 129–43.

Bowen, H. (2011). Spelling it out! Accounting for spelling difficulties for Arab learners of English. *HCT Ma'rifa.* UAE: Higher Colleges of Technology.

Bowles, M. A. (2008). Task type and reactivity of verbal reports in SLA: A first look at an L2 task other than reading. *Studies in Second Language Acquisition, 30*, 359-387.

Brevik, L., Olsen, R., & Hellekjær, G. (2016). The complexity of second language reading: Investigating the relationship between L1 and L2 using upper secondary level national reading tests. *Reading in a Foreign Language, 28*,161-182.

Browne, C. Culligan, B. & Phillips, J. (2013). The New General Service List. http://www.newgeneralservicelist.org/

Brunfaut, T. (2016*). Looking into reading: A follow-up study on test-takers' cognitive processes while completing APTIS B1 reading tasks*. English Language Assessment Research Group, London, UK: British Council.

Brunfaut, T. & McCray, G. (2015). *Looking into test-takers' cognitive processes whilst completing reading tasks: A mixed-method eye-tracking and stimulated recall study*. (ARAGs Research Reports Online; Vol. AR/2015/001). London, UK: British Council.

Buswell, G. (1922). *Fundamental reading habits: A study of their development.* Chicago, IL: University of Chicago Press.

Buswell, G. (1937). How adults read. *Supplementary Educational Monograph, 45*. Chicago, IL: University of Chicago Press.

Cain, K. (2006). Children's reading comprehension: The role of working memory in normal and impaired development. In S. Pickering (Ed.), *Working memory in education* (pp. 61-91). Burlington: MA: Academic Press.

Cain, K. & Oakhill, J. (2006) Profiles of children with specific comprehension difficulties. *British Journal of Educational Psychology, 76*(4), 683-696.

Cain, K. & Oakhill, J. (2014) Reading comprehension and vocabulary: is vocabulary more important for some aspects of comprehension? *Annee Psychologique, 114*(4), 647-662.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability and component skills. *Journal of Educational Psychology*, *96*, 31-42.

Carr, T. & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T., Waller, & G. MacKinnon (Eds.), *Reading research: Advances in theory and practice*, (Vol. 5, pp. 1-82). Orlando, FL: Academic Press.

Carreiras, M. & Clifton, C. (2004). On the on-line study of language comprehension. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP and beyond.* (pp.1-14). Brighton, UK: Psychology Press.

Cataldo, M. & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in a text. *Journal of Educational Psychology, 92*(4),791-799.

Chase, K., Rayner, K. &Well, A. (2005). Eye movements and phonological parafoveal preview: Effects of reading skill. *Canadian Journal of Experimental Psychology, 59*, 209-217.

Chiappe, P., Glaeser, B. & Ferko, D. (2007). Speech perception, vocabulary, and the development of reading skills in English among Korean- and English-speaking children. *Journal of Educational Psychology*, *99*, 154–166.

Christie, J. & Just, M. (1976). Remembering the location and content of sentences in a prose passage. *Journal of Educational Psychology. 68*, 702–710.

Clarke, M. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning, 29*, 121-150.

Clarke, M. (1980). The short circuit hypothesis of ESL reading-or when language competence interferes with reading performance. *The Modern Language Journal, 64*, 203-209.

Clarke, M. (1994). The dysfunction of the theory/practice discourse. *TESOL Quarterly, 28*, 9-26.

Clifton, C., Ferreira, F., Henderson, J., Inhoff, A., Liversedge, S., Reichle, E. & Schotter, E. (2016). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language, 86,* 1-19.

Clifton, C., Staub, A & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel, M. Fischer, W. Murray & R. Hill (Eds.*), Eye Movements: A Window on Mind and Brain* (pp. 341-372). New York, NY: Elsevier.

Coady J. (1979). A psycholinguistic model of the ESL reader. In R. Mackay B. Barkman & R. Jordan (Eds.), *Reading in a Second Language* (pp. 5-12) Rowley, MA: Newbury House Publishers.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London, UK: Academic Press.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education, 4*, 5-17.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589-608.

Coltheart, M, Rastle, K. & Perry, C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Conklin, K. & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 3*, 453-476.

Cowan, J. & Sarmad, Z. (1976). Reading performance of bilingual children according to type of school and home language. *Language Learning, 26*, 353-376.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessmen*t. Cambridge, U.K: Press Syndicate of the University of Cambridge.

Cowie, R., Douglas-Cowie, E. & Wichmann, A. (2002). Prosodic characteristics of skilled reading: Fluency and expressiveness in 8–10-year-old readers. *Language and Speech, 45*, 47–82.

Crookes, G. (1993). Action research for second language teachers: Going beyond teacher research. *Applied Linguistics, 14*,130-144.

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

Crossley, S., Greenfield, J. & McNamara, D. (2008). Assessing text readability using cognitively-based indices. *TESOL Quarterly, 42*, 475-493.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism, 9,* 1-43.

Cummins, J. (1979) Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19,* 121-129.

Cunningham, A. (2005). Vocabulary growth through independent reading and children reading aloud to. In E. Hiebert and M. Kamil (Eds.*), Teaching and learning vocabulary: Bringing research to practice* (pp. 45-68). Mahwah, NJ: Erlbaum.

Cunningham, A., Perry, K., Stanovitch, K. & Stanovitch, P. (2004). Disciplinary knowledge of K-3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia, 54,* 39-167.

Cunningham, A. & Ryan O'Donnell, C. (2015). Teacher knowledge in early literacy. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading (*pp. 447-462). New York, NY: Oxford University Press.

Cunningham, A. & Zibulsky, J. (2009). Perspectives on teachers' disciplinary knowledge of reading processes, development and pedagogy. *Reading & Writing: An Interdisciplinary Journal, 22,* 375-378.

Daneman, M, Carpenter, P. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *1*9, 450–466.

Daneman, M. & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422–433.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-151). New York, NY: Cambridge University Press.

Deutsch, A., Frost, R., Pelleg, S., Pollatsek, D. & Rayner, K. (2003). Early morphological effects in reading: Evidence from parafoveal preview benefit in Hebrew. *Psychonomic Bulletin & Review, 3,* 415-422.

Dörnyei, *Z*. (2009). *The psychology of second language acquisition*. Oxford, UK: Oxford University Press.

Dowhower, S. (1991). Speaking of prosody: Fluency's unattended bedfellow [Electronic version]. *Theory into Practice, 30*, 165-175.

Dussias, P. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, *30*,149-166.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London, UK: Springer.

Egi, T., Adams, R. & and Neuvo, A. (2013). Is metalinguistic stimulated recall reactive in second language learning? In J. Bergsleithner, S. Frota & J. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 81-102). Honolulu: National Foreign Language Resource Centre.

Ehri, L. (1980). The development of orthographic images. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 311–338). London, UK: Academic Press.

Ehri, L (1984). How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), *Language awareness and learning to read* (pp. 119-147). New York, NY: Springer Verlag.

Ehri, L. (1987). Learning to read and spell words. *Journal of Reading Behavior Volume XIX*, 1, 5-31.

Ehri, L. (1989). The development of spelling knowledge and its role in reading acquisition and reading disability. *Journal of Learning Disabilities, 22,* 356-365.

Ehri, L. (1991). Learning to read and spell words. In L. Rieben & C. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 57-73). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ehri, L. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri and R. Treiman (Eds.*), Reading acquisition* (pp. 107-143). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ehri, L (1997) Learning to read and learning to spell are one and the same, almost. In C. Perfetti, A. Rieben, M. Fayol, (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 237-269). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Ehri, L. (1998). Word reading by sight and by analogy in beginning readers. In C. Hulme and R. Joshi (Eds.), *Reading and spelling; Development and disorders* (pp. 87-111). Mahwah, NJ: Lawrence Erlbaum Associates.

Ehri, L. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167–188.

Ehri, L, Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z. & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the

National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287.

Ehri, L. & Rosenthal, J. (2007). Spellings of Words: A neglected facilitator of vocabulary learning, *Journal of Literacy Research, 39,* 389-409.

Ehri, L. & Snowling, M. (2004). Developmental variation in word recognition. In C. Stone, C. Silliman, E. Ehren & K. Apel, K. (Eds.), *Handbook of language and literacy: Development and disorder (*pp. 433-460). New York, NY: Guilford.

Ehri, L. & Wilce, L. (1979). The mnemonic value of orthography among beginning readers. *Journal of Educational Psychology, 71,* 26-40.

Ehri, L. & Wilce, L. (1980). The influence of orthography on readers' conceptualization of the phoneme structure of words. *Applied Psycholinguistics, 1,* 371-385.

Eklundh, K. S. (1992). Problems in achieving a global perspective of the text in computer-based writing. *Instructional Science, 21,* 73–84.

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition, 18,* 91-126.

Ellis, N. (2012b). Formulaic language and the phrasal teddy bear. Annual Review of Applied Linguistics, 32, 17-44.

Ellis, N. & Cadierno, T. (2009).Constructing a second language. Introduction to the Special Section. *Annual Review of Cognitive Linguistics, 7,* 111-139.

Ellis, R. (1997a). SLA and language pedagogy: An educational perspective. *Studies in Second Language Acquisition, 19,* 69–92.

Ellis, R. (2001). Second language acquisition: Research and language pedagogy. In C. Candlin & N. Mercer (Eds.), *English language teaching in its social context* (pp. 44–74). London, UK: Routledge

Ehmke, C. & Wilson, S. (2007). Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on HC*, *1*, (pp. 119-128). Swinton, UK: British Computer Society.

Erman, B. &Warren, B. (2000). The idiom principle and the open-choice principle. *Interdisciplinary Journal for the Study of Discourse, 20,* 29-62.

Farid, M. & Grainger, J. (1996). How initial fixation position influences visual word recognition: A Comparison of French and Arabic. *Brian and Language, 53,* 351-368.

Fender, M. (2003). English word recognition and word integration skills of native Arabic and Japanese-speaking learners of English. *Applied Psycholinguistics*, *24*, 289-315.

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language*, *20*, 19-42.

Field, J. (2004). *Psycholinguistics: The key concepts*. London, UK: Routledge.

Figueredo, L. (2006). Using the known to chart the unknown: A review of first-language influence on the development of English-as-a-second-language spelling skill. *Reading and Writing: An Interdisciplinary Journal, 19,* 873–905.

Findley, J. & Gilchrest, I. (2008). *Active vision: The psychology of looking and seeing.* London: Oxford University Press.

Foorman, B., Francis, D., Fletcher, J., Schatschneider, C. & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading disabilities in at-risk children. *Journal of Educational Psychology, 90,* 37–55.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,*178-210.

Fukkink, R., Hulstijn, J. & Simis, A. (2005). Does training in second language word recognition skills affect reading comprehension? An experimental study. The *Modern Language Journal, 89,* 54-75.

Garcia, J. & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*, 74-111.

General Secretariat for Development and Planning (2008). *Qatar National Vision 2030*. Doha: Qatar.

Geva, E. & Siegel, L. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal, 12*, 1–30.

Geva, E. & Zadeh, Z. (2006). Reading efficiency in native English-speaking and English-as-a-second-language children: The role of oral proficiency and underlying cognitive-linguistic processes. *Scientific Studies of Reading*, *10*, 31–57.

Godfroid, A., Boers, F. & Housen, A. (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.

Godfroid, A., Housen, A. & Boers, F. (2010). A procedure for testing the noticing hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Cognitive*

processing in second language acquisition (pp. 169-197). Amsterdam, NL: John Benjamins.

Godfroid, A., Loewen, S., Jung, S., Park, J. H., Gass, S. & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*, 269-297.

Godfroid, A. & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye movement data, verbal reports and vocabulary learning scores. In J. Bergsleithner, S. Frota & J. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183-205). Hawaii: National Foreign Language Resource Center.

Godfroid, A. & Uggen, M. (2013). Attention to irregular verbs by beginning learners of German. *Studies in Second Language Acquisition, 35*, 291-322.

Godfroid, A. & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye movement data. In P. Rebuchat (Ed.), *Recording eye movements in second language research* (pp. 325-348). New York, NY: Routledge.

Goodman, K. (1970). Reading as a psycholinguistic guessing game. In H. Singer & R. Ruddell (Eds.), *Theoretical models and Processes of Reading* (pp. 259-271). Newark, N.J.: International reading Association.

Goodman, K. (1973) Miscues: Windows on the reading process. In K. Goodman (Ed.), *Miscue analysis: Applications to reading instruction* (pp. 3-14). Urbana, IL: NCTE.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, *42*, 73– 83.

Gough, P. & Tunmer, W. (1986). Decoding, reading and reading disability. *Remedial and Special Education, 7*, 6-10.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly, 25*, 375-406.

Grabe, W. (2009*). Reading in a second language: Moving from theory to practice.* New York, NY: Cambridge University Press.

Grabe, W. (2010). Fluency in reading-thirty-five years later. *Reading in a Foreign Language, 2,* 71-83.

Grabe, W. (2012). Reading in a second language. In *Oxford handbook of applied linguistics*. New York, NY: Oxford University Press.

Graesser, A., McNamara, D. & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40,* 223-234.

Graesser, A., McNamara, D., Louwerse, M. & Cai, Z. (2004). Coh-Metrix: Analysis of text cohesion and language. *Behavior Research Methods, Instruments & Computers, 36,* 193-202.

Graham, S., Harris, K. & Chorzempa, B. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology, 94*, 669-686.

Granger, S. & Meunier, F. (Eds.), (2008) *Phraseology. An interdisciplinary perspective*. Amsterdam, NL: John Benjamins.

Gray, C. (1917). Types of reading ability as exhibited through tests and laboratory experiments. *Supplementary Educational Monograph, 1*, (5).

Hacquebord, H. (1989). *Tekstbegrip van Turkse en Nederlandse leerlingen in het voortgezet onderwijs [Reading comprehension of Turkish and Dutch students attending secondary schools]*. Groningen, NL: RUG.

Hama, M. & Leow, R. (2010). Learning without awareness revisited. Extending Williams (2005). *Studies in Second Language Acquisition, 32*, 465-401.

Hamada, M. & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58,* 1-31.

Han, Z., Park, E. & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics, 29*, 597-618.

Hanks, J. (2015). Language teachers making sense of Exploratory practice. *Language Teaching Research, 19,* 612-633.

Hannon, B. & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93,* 103-128.

Harding, L., Alderson, C. & Brunfaut, T. (2015). Diagnostic assessment of reading and listening in a second or foreign language: Elaborating on diagnostic principles, *Language Testing, 32*, 317–336.

Harrison, G. & Krol, L. (2007). The relationship between L1 and L2 word-level reading and phonological processing in ESL adults. *Journal of Research in Reading, 30*, 379-393.

Hayes-Harb, R. (2006). Native speakers of Arabic and ESL texts: Evidence for the transfer of written word identification processes. *TESOL Quarterly, 40,* 321-339.

Hiver, P. & Dörnyei, Z. (2015). Language teacher immunity: A double-edged sword. *Applied Linguistics, 38*, 405-423

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodkka, H. & Van De Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. New York, NY: Oxford University Press.

Hoover, W. & Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Howard, A. (2002). The relationships of phonemic awareness and rapid naming to reading development, *Learning Disability Quarterly 25*, 47-57.

Huey, E. (1908). *The psychology and pedagogy of reading*. New York, NY: Macmillan.

Hutzler, F. & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language, 89,* 235-242.

Hyönä, J. & Olsen, R. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology Learning, Memory Cognition, 21*(6), 1430-40.

Indrarathne, B. & Kormos, J. (2016). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition, 27*, 1-30.

Inhoff A., Greenberg, S., Solomon, M. & Wang, C. (2009). Word integration and regression programming during reading: A test of the E-Z Reader 10 model. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1571-1584.

Inhoff, A. & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception (*pp. 29- 54). Amsterdam, NL: Elsevier.

Inhoff, A. & Rayner, K. (1986). Parafoveal word processing during fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*, 431-439.

Izumi, S. (2002). Output enhancement and the noticing hypothesis. *Studies in Second Language Acquisition, 24,* 541-577.

Jahan, A. & Kormos, J. (2014). The impact of textual enhancement on EFL learners' grammatical awareness of future plans and intentions *International Journal of Applied Linguistics, 25,* 46-66.

Jegerski, J. & Van Patten, B. (2013). *Research methods in second language psycholinguistics*. New York, NY: Routledge.

Jiang, X. (2016). The role of oral reading fluency in ESL reading comprehension among learners of different first language backgrounds. *The Reading Matrix: An International Online Journal, 16,* 227-242.

Jiang, X., Sawaki, Y. & Sabatini, J. (2012). Word reading efficiency and oral reading fluency in ESL reading comprehension. *Reading Psychology, 33*, 323–349.

Johnson-Laird, P. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.

Joseph, H., Liversedge, S., Blythe, H., White, S. & Rayner, K. (2008). Word length and landing position effects during reading in children and adults. *Vision Research, 49,* 2078-2086.

Jordan, T., Almabruk, A., Gadalla, E., McGowan, V., White, S., Abedipour, L. & Paterson K. (2014). Reading direction and the central perceptual span: Evidence from Arabic and English. *Psychonomic Bulletin & Review, 21,* 505-511.

Judd, C. & Buswell, G. (1922). *Silent reading: A study of various types*. Chicago, IL: University of Chicago Press

Juhasz, B. & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology, Learning Memory and Cognition, 29,* 1312-1318.

Juhasz, B. & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition, 13,* 846-863.

Just, M. & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Katz, L. & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, Phonology, Morphology, and Meaning* (pp. 67-84). Amsterdam, NL: Elsevier North Holland Press.

Kelman, M. & Apel, K. (2004). The effects of a multiple linguistic, prescriptive approach to spelling instruction: A case study. *Communication Disorders Quarterly, 25*(2), 56-66.

Kennedy, A., Brooks, R., Flynn L-A & Prophet, C. (2003). The reader's spatial code. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 193-212) Amsterdam, NL: Elsevier.

Kennedy, A. & Murray, W. (1987). Spatial coordinates and reading: Comments on Monk (1985), *Quarterly Journal of Experimental Psychology, 39A,* 649–656.

Khalifa, H. & Weir, C. (2009). *Examining reading: Research and practice in assessing second language reading.* Cambridge,UK: UCLES/Cambridge University Press.

Kincaid, J., Fishburne, J., Robert P., Rogers, R. & Chissom, B. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.* Naval Technical Training Command, Millington TN: Research Branch.

Kirkby, J., Webster, L., Blythe, H. & Liversedge, S. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin, 134,* 742-763.

Klauda, S. & Guthrie, J. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100*, 310–321.

Koda, K. (2004). *Insights into second language reading*. New York, NY: Cambridge University Press.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57,* 1-44.

Kuhn, M., Schwanenflugel, P. & Meisinger, E. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*, 230-251.

Kuhn, M. & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *The Journal of Educational Psychology, 95,* 3-21.

Laberge, D. & Samuels, S. (1974*).* Towards a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Lambert, W. (1975). Culture and language as factors in learning and education. In A.Wolfgang (Ed.), *Education of Immigrant Students* (pp. 55-83). Toronto, CA: O.I.S.E.

Lamme, V. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*, 12–18.

Laufer, B. & Sim, D. (1985). Measuring and explaining the reading threshold needed for English for academic purposes texts. *Foreign Language Annals, 18*(5), 405-411.

Larsen-Freeman, D. (1998). Learning teaching is a lifelong process. *Perspectives xxiv,2*,5-11.

Larsen-Freeman, D. & Long, M. (1991). *An introduction to second language acquisition research.* New York, NY: Longman.

Lee, S. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning, 57,* 87-118.

Lefton, L., Nagle, R., Johnson, G. & Fisher, D. (1979). Eye movement dynamics of good and poor readers: Then and now. *Journal of Reading Behaviour, 11,* 319–328.

Legge, G., Klitz, T.& Tjan, B. (1997). Mr. Chips: An ideal-observer model of reading. *Psychological Review, 104,* 524-53.

Leow, R. (1997). Attention, awareness, and foreign language behaviour. *Language Learning,* 47, 467-505.

Leow, R. (2000). A study of the role of awareness in foreign language behaviour. *Studies in Second Language Acquisition, 22,* 557-584.

Leow, R. & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition, 26,* 35-57.

Lesaux, N. & Siegel, L. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39,* 1005-1019.

Levy-Schoen, A. (1981). Flexible and/or rigid control of oculomotor scanning behavior. In D. Fisher, R. Monty & J. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 299-314). Hillsdale, N.J.: Erlbaum Associates.

Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation, 61*(6), 700-712.

Liversedge, S., Paterson, K. & Pickering, M. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), Eye movements and measures of reading, (pp.55-76). Cambridge, MA: Springer.

Loftus, G. & Mackworth, N. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565-572.

Long, M. (2015). *Second language acquisition and task-based language teaching*. Hoboken, NJ: Wiley.

Longcamp, M., Boucard, C., Gilhodes, J.C., Anton, J.L., Roth, M., Nazarian, B. & Velay, J-L. (2008). Learning through hand or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience, 20,* 802-815.

Lueck, K.,Mendez, M. & Perryman, K. (2000). Eye movement abnormalities during reading in patients with Alzheimer disease. *Cognitive and Behavioural Neurology, 13,* 77-82.

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27,* 405-430.

MacNamara, J. (1970). Bilingualism and thought. In J. Alatis (Ed.), *Bilingualism and language contact: Anthropological, linguistic, psychological and sociological aspects* (pp. 5-40). Georgetown Round Table of Languages and Linguistics, Washington, DC: Georgetown University.

Mangen, A., Anda, L., Oxborough, G. & Brønnick, K. (2015). *Journal of Writing Research, 7,* 227-247.

Mangen, A., Robinet, P., Olivier, G. & Velay, J.L. (2014, July 21-25). *Mystery story reading in pocket print book and on Kindle: possible impact on chronological events memory*. Paper presented at IGEL (The International Society for the Empirical Study of Literature and Media), Turin, Italy.

Mangen, A., Walgermo, B. & Brønnick, K. (2013). Reading linear texts on paper vs. computer screen: Effects on reading comprehension. *International Journal of Educational Research, 58,* 61-68.

Martinez, R. & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33,* 299-320.

Masterson, J. & Apel, K. (2000). Spelling assessment: Charting a path to optimal instruction. *Topics in Language Disorders, 20*, 50-6.

Masterson, J. & Apel, K. (2007). Spelling and word-level reading: A multilinguistic approach. In A. Kamhi, J. Masterson & K. Apel (Eds.), *Clinical decision making in developmental language disorders* (pp. 249–266). Baltimore, MD: Paul H. Brookes Publishing.

Masterson, J. & Apel, K. (2010). The spelling sensitivity score: Noting the developmental changes in spelling. *Assessment for Effective Intervention, 36*, 35-45.

Mather, N., Bos, C. & Babur, N. (2001). Perceptions and knowledge of pre-service and in-service teachers about early literacy instruction. *Journal of Learning Disabilities, 34*, 472‑482.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception. Part I: An account of basic findings. *Psychological Review, 88*, 375-407.

McClelland, J. & Rumelhart, D. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

McConkie, G., Kerr, P., Reddix, M. & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research, 28,* 1107-1118.

McConkie, G., Kerr, P., Reddix, M., Zola, D. & Jacobs, A. (1989). Eye movement control during reading: Frequency of refixating a word. *Perception & Psychophysics, 46*, 245-253.

McConkie, G. & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics, 17*, 578-586.

McConkie, G. & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society, 8,* 365-368.

McConkie, G., Zola, D., Grimes, J., Kerr, P., Bryant, N. & Wolff, P. (1991). Children's eye movements during reading. In J. Stein (Ed.), *Vision and visual dyslexia* (pp. 251-262). Boston, M.A.: CRC Press.

McNamara, D. (2013). The epistemic stance between the author and the reader: A driving force in the cohesion of text and writing. *Discourse Studies, 15*(5), 1-17.

McNamara, D. & Graesser, A. (2012). Coh-Metrix: An automated tool for theoretical and applied natural processing. In P. McCarthy & C. Boonthum (Eds.), *Applied natural language processing: Identification, investigation and resolution* (pp. 188-205). Hershey, P.A.: IPA Global.

McNamara, D., Graesser, A., McCarthy, P. & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, M.A.: Cambridge University Press

Meseguer, E., Carreiras, M., & Clifton, C. Jr. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition, 30*, 551-562.

Miellet, S. & Sparrow, L. (2004). Phonological codes are assembled before word fixation: Evidence from boundary paradigm in sentence reading. *Brain and Language, 90*, 299-310.

Moats, L. (1994). The missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia, 44*, 81–102.

Moats, L. (2000). *Speech to print: Language essentials for teachers*. Baltimore, MD: Paul H. Brookes.

Morris, R., Rayner, K. & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 268-281.

Morris, D., Trathan, W., Frye, E., Kucan, F., Ward, D. & Schlagal, R. (2012). The role of reading rate in the informal assessment of reading ability. *Literacy Research and Instruction, 52*, 52-64.

Mueller, P. & Oppenheimer, D. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note-taking. *Psychological Science 23*(6), 1159-1168.

Murray, W. & Kennedy, A. (1998). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analysis. *The Quarterly Journal of Experimental Psychology, 40,* 693-718.

Nagy,W. & Scott, J. (2000). Vocabulary processes. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research, Vol. III* (pp. 269–284). Mahwah, NJ: Erlbaum.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal, 87,261*-276.

Nassaji, H. (2012). The relationship between SLA research and pedagogy: Teachers' perspectives. *Language Teaching Research, 16*, 337-365.

Nassaji, H. & Fotos, S. (2011). *Teaching grammar in second language classrooms. Integrating form-focused instruction in communicative context*. London, U.K.: Routledge.

Nassaji, H. & Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics, 20*, 241–267.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from http://www.nichd.nih.gov/publications/nrp/report.cfm.

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology, 54,* 31-39

Oakhill, J., Cain, K. & Bryant, P. (2003). The dissociation of single-word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18,* 443-468.

Oakhill, J., Cain, K.& Elbro, C. (2015). *Understanding and teaching reading comprehension. A Handbook.* New York, NY: Routledge.

O'Brian, J. (1922). *Silent reading: With special reference to methods for developing speed: A study in the psychology and pedagogy of reading*. New York, NY: Macmillan.

O'Brien, B., Wolf, M., Miller, L., Lovett, M. & Morris, R. (2011). Orthographic processing efficiency in developmental dyslexia: An investigation of age and treatment factors at the sublexical level. *Annals of Dyslexia, 61*, 111–135.

O'Regan, J. (1981). The convenient viewing position hypothesis. In D. Fisher, R. Monty, & J. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 289-298). Hillsdale, NJ: Erlbaum.

O'Regan, J. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 333-354), New York, NY: Springer-Verlag.

O'Regan, J. & Levy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 363-383). Hillsdale, NJ: Erlbaum.

Paulson, E. & Goodman, K. (1999, January). Influential studies in eye movement research. *Reading Online*. Retrieved from http://www.readingonline.org/research/eyemove.html

Pellicer- Sánchez, A. (2012). Vocabulary and reading. In C. Chapelle (Ed.), *Encyclopedia of applied linguistics.* Wiley Blackwell.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*, 97-130.

Pellicer-Sánchez, A. & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language, 22*, 31-55.

Perfetti, C. (1984). Reading acquisition and beyond: Decoding includes cognition. *American Journal of Education, 93*, 40-60.

Perfetti, C. (1985). *Reading Ability*. New York, NY: Oxford University Press.

Perfetti, C. (1988). Verbal efficiency in reading ability. In M. Daneman, G. Mackinnon, & T. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 6, pp. 109-143). New York, NY: Academic Press.

Perfetti, C. (1992). The representation problem in reading acquisition. In P. Gough, L. Ehri & R. Treiman (Eds.), *Reading Acquisition* (pp. 145-174*).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. (1997). The psycholinguistics of spelling and reading. In C. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice* (pp. 21-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Perfetti, C. & Lesgold, A. (1977). Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 41-183). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C., Rieben, L. & Fayol, M. (1997). *Learning to spell: Research, theory, and practice across languages*. Mahwah, NJ: Lawrence Erlbaum Associates.

Perfetti, C. & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22-37.

Pikulski, J. & Chard, D. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510-519.

Piolat, A., Roussey, J.Y. & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human–Computer Studies, 47,* 565–589.

Plummer, P. & Rayner, K. (2012). Effects of parafoveal word length and orthographic features on initial fixation landing positions in reading. *Attention, Perception & Psychophysics, 74*, 950-963.

Pollatsek, A. & Rayner, K. (1982). Eye movement control during reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance 8*, 817-833.

Pollatsek, A., Bolozky, S., Well, A. & Rayner, K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language, 14*, 174-180.

Prado C., Dubois, M. & Valdois, S. (2007). The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span. *Vision Research, 47*(19), 2521-2530.

Pressey, S. (1926). A simple apparatus which gives tests and scores and teaches. *School and Society, 25*, 549-552.

Pressley, M. (2000). What should reading instruction be the instruction of? In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of Reading Research: Volume 3* (pp. 545-561). New York, NY: Longman.

Pynte, J. & Kennedy, A. (2006). An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research, 46*, 3786-3801.

Qian, D. (1999). Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review, 56*, 282–307.

R Core Team. (2012). R: A language and environment for statistical computing [Computer Software]. Retrieved from R Foundation for Statistical Computing http://www.R-project.org/.

Randall, M. (2007). *Memory, psychology and second language learning*. Amsterdam, NL: John Benjamins.

Randall, M. & Groom, N. (2009). The BUiD Arab learner corpus: A resource for studying the acquisition of spelling. Retrieved from http://ucrel.lancs.ac.uk/publications/CL 2009/54_FullPaper.doc.

Randall, M. & Meara, P. (1988). How Arabs read Roman letters. *Reading in a Foreign Language, 4,* 133-145.

Rasinski, T. (2006). Reading fluency instruction: Moving beyond accuracy, automaticity, and prosody. *The Reading Teacher, 59,* 704-706.

Rasinski, T. (2010). *The fluent reader*. New York, NY: Scholastic.

Rasinski, T. (2012). Why reading fluency should be hot. *The Reading Teacher, 65,* 516–522.

Rasinski, T. & Hoffman, J. (2003). Oral reading in the school reading curriculum. *Reading Research Quarterly, 38,* 510–522.

Rasinski, T., Rikli, A. & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48,* 350–361.

Rasinski, T. & Samuels, S. (2011). Reading fluency: What it is and what it is not. In S. Samuels & A. Farstrup (Eds.), *What research has to say about reading instruction* (pp. 94-113). Newark, DE: International Reading Association.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65-81.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85,* 618-660.

Rayner, K. (1979). Eye guidance in reading. Fixation locations in words. *Perception, 8,* 21-30.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41,* 211-236.

Rayner, K. (1988). Eye movements in reading and information processing: 20 years of research. *Psycholinguistic Bulletin, 124,* 372-422.

Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading, 1,* 317-339.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1502.

Rayner, K. (2012). *Children of the Code Interview*. Retrieved from http://www.childrenofthecode.org/interviews/Rayner.htm.

Rayner, K. & Bertera, J. (1979). Reading without a fovea. *Science, 206,* 468-469.

Rayner, K., Chace, K., Slattery, T. & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10,* 241-255.

Rayner, K. & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity and lexical ambiguity. *Memory & Cognition, 14,* 191-201.

Rayner, K. & Fischer, M. (1996). Mindless reading revisited. Eye movements during reading and scanning are different. Perception and *Cognition, 58,* 734-747.

Rayner, K., Foorman B., Pesetsky, D. & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *American Psychological Society, 2,* 31-74.

Rayner, K., Juhasz, B. & Brown, S. (2007). Do readers obtain preview benefit from word n+2? A test of serial attention shift versus distributed lexical processing of eye movement control in reading. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 230-245.

Rayner, K. & McConkie, G. (1976). What guides a reader's eye movements? *Vision Research, 16,* 829-837.

Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., Pollatsek, A. & Binder, K. (1998). Phonological codes and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 476-497.

Rayner K., Reichle, E., Stroud, M., Williams, C. & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and elderly readers. *Psychology & Aging, 21*, 448-465.

Rayner, K., & Sereno, S. (1994). Eye movements in reading: Psychological studies. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic.

Rayner, K., Sereno, S. & Raney, G. (1996). Eye movement control in reading. A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1188-12.00.

Rayner, K., Slattery, T., Dreighe, D. & Liversedge, S. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Human Perception and Performance, 37*, 514-528

Rayner, K., Slowiaczek, M., Clifton, C. & Bertera, J. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 912-922.

Rayner, K., Well, A. & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics, 27*, 537-544.

Reichle, E. Rayner, K. & Pollatsek, A. (2003). The E-Z reader model of eye movement control in reading: comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.

Reichle, E. Warren, T. & McConnell, K. (2009). Using E-Z reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*(1), 1-20.

Reilly, R. & O'Regan, J. (1998). Eye movement control in reading: A simulation of some word-targeting strategies. *Vision Research, 38,* 303-317

Rumelhart, D., Hinton, G. & Williams, R. (1986). Learning internal representations by error propagation. *Nature,323*,533-536.

Rigg, P. (1977) 'The miscue ESL project.' In H. Brown, C. Yorio, & R. Crymes (Eds.), *On TESOL '77: Teaching and learning ESL: Trends in research and practice* (pp. 106-181). Washington, DC: TESOL.

Robinson, P. (1995). Review article: Attention, memory and the noticing hypothesis. *Language Learning, 45,* 283-331.

Ryan, A. (1997). Learning the orthographical form of L2 vocabulary-a receptive and a productive process. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 181-198). Cambridge, UK: Cambridge University Press.

Ryan, A. & Meara P. (1991). The case of the invisible vowels: Arabic speakers reading English words. *Reading in a Foreign Language, 7,* 531-540.

Sachs, R. & Polio, C. (2007). Learners' uses of two types of written feedback on an L2 writing revision task. *Studies in Second Language Acquisition, 29,* 67-100.

Sadhwani, P. (2005). *Phonological and orthographic knowledge: An Arab-Emirati perspective* (Unpublished master's dissertation). British University in Dubai, Dubai: United Arab Emirates.

Saigh, K. & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System, 40*, 24-36.

Salvucci, D. & Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the eye tracking research and applications symposium*, 71-78. New York, NY: ACM Press

Samuels, S., Rasinski, T. & Hiebert, E. (2011). Eye movements and reading: What teachers need to know. In A. Farstrup & S. Samuels (Eds.), *What research has to say about reading instruction* (pp. 25-50) Newark, DE: IRA.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness on learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Hawaii: University of Hawaii.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. (2010 December 2-4). Attention, awareness, and individual differences in language learning. In W. Chan, S. Chi, K. Cin, J. Istanto, M. Nagami, J. Sew, T. Suthiwan & I. Walker (Eds.), *Proceedings of CLaSIC 2010*, Singapore, (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

Schmidt, W. (1917). An experimental study in the psychology of reading. A dissertation. *Supplementary Educational Monographs, 1*(2). Chicago, IL: The University of Chicago Press.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. London, UK: Palgrave McMillan.

Schoonen, R., Hulstjin, J. & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among dutch students in grades 6, 8 and 10. *Language Learning, 48*, 71-106.

Schotter, E., Angele, R., & Rayner, K. (2006). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, *74*, 5-35.

Schreiber, P. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice, 30*, 158-164.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 382-408). Malden, MA: Blackwell Publishing.

Segalowitz, N. & Segalowitz, S. (1993). Skilled performance, practice and the differentiation of speedup from automatization effects. *Applied Linguistics, 14*, 369–385.

Sereno, S., O'Donnell, P. & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate-bias effect. *Journal of Experimental Psychology: Human Perception and performance, 32*, 335-350.

Sereno, S. & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Science, 7*, 489–493.

Share, D. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–218.

Sharwood Smith, M (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research, 7*, 118-132.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition, 15,* 165–179.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Siegler, R. (1996a). *Emerging minds: The process of change in children's minds*. New York, NY: Oxford University Press.

Simard, D. (2009). Differential effects of textual enhancement formats on intake. *System, 37*, 124-135.

Simard, D. & Foucambert, D. (2013). Observing noticing while reading in L2. In J. Bergsleithner & S. Frota (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 215-234). Honolulu, HI: University of Hawai`i, National Foreign Language Resource Center.

Sinclair, J. (2005, October). *The phrase, the whole phrase and nothing but the phrase* [Unpublished plenary lecture]. Phraseology 2005, Louvain-la-Neuve, Belgium.

Sirkin, R. (2006). *Statistics for the social sciences* (3rd Ed.). Thousand Oaks, CA: Sage.

Smith, B. (2012). Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning & Technology, 16*, 53-81.

Smith, B. & Renaud C. (2013). Using eye tracking as a measure of foreign language learners' noticing of recasts during computer-mediated writing conferences. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (pp. 147-165). Amsterdam,NL: John Benjamins.

Smoker, T., Murphy, C. & Rockwell, A. (2009). Comparing memory for handwriting versus typing. Paper presented at the Human Factors and Ergonomics Society Annual Meeting. http://dx.doi.org/10.1177/154193120905302218

Snow, C, Burns, S. & Griffin, P. (Eds.), (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Spada, N. & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 1-46.

Spinner, P., Gass, S. & Behney, J. (2013). Coming eye-to-eye with noticing. In J. Bergsleithner, S. Frota, & J. Yoshioka, (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 235–254). Honolulu, HI: University of Hawai ‘i, National Foreign Language Resource Center.

Stahl, S. & Kuhn, M. (2002). Making it sound like language: Developing fluency. *The Reading Teacher, 55,* 582-584.

The National Staff Development Council & the School Redesign Network. (2009). *Status Report on Teacher Development in the United States and Abroad: Professional learning in the learning profession*. Stanford, CA: Stanford University.

Taha, H. (2013) Reading and Spelling in Arabic: Linguistic and Orthographic Complexity *Theory and Practice in Language Studies, 3,* 721-727.

Taylor, S. (1971). The dynamic activity of reading: A model of the process. *EDL Research and Information Bulletin,* (Vol. 9). New York, NY: McGraw-Hill.

Thompson-Panos, K. &Thomas-Ružič, M. (1983). The least you should know about Arabic: Implications for the ESL writing instructor. *TESOL Quarterly*, *17,* 609-623.

Tobii Technology (2010). How are different heat maps calculated? [White paper]. Retrieved from Tobii website: http://www.tobii.com/en/eye-tracking-research/global/library/white-papers/tobii-eye-tracking-white-paper/

Tobii Technology (2014). Tobii Studio™ User Manual. Danderyd, Sweden:Tobii AB.

Tomlin, R. & Villa, V. (1994). Attention in cognitive science and SLA. *Studies in Second Language Acquisition, 16*,183-203.

Torgesen, J., Rashotte, C. & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 333–355). Timonium, MD: York Press.

Tunmer, W. & Chapman, J. (1998). Language prediction skill, phonological recoding ability, and beginning reading. In C. Hulme, & R. Joshi (Eds.), *Reading and spelling: Development and Disorder* (pp. 33–67). Hillsdale, NJ: Erlbaum.

Urquhart, A. & Weir, C. (1998). *Reading in a foreign language*. London, UK: Longman.

vanGelderen, A., Schoonen, R., Stoel, R., deGlopper, K & Hulstjin, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology, 99,* 477-491.

van Gelderen, A., Schoonen, R.,de Glopper, K., Hulstijn, J., Simis, A., Snellings, P.& Stevenson, M. (2004). *Journal of Educational Psychology, 96,* 19-30.

Veenendaal, N. Groen, M. &Verhoeven, L. (2015). What oral text reading can reveal about reading comprehension. *Journal of Research in Reading, 38(3),* 213-225.

Vitu, F. (1991). The influence of parafoveal processing and linguistic control on the optimal landing position effect. *Perception & Psychophysics, 50*, 58-75.

Vitu, F. & McConkie, G. (2000). Regressive saccades and word perception in adult reading. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 301–326). Amsterdam, NL: Elsevier.

Vitu, F. & O'Regan, J. (1995). A challenge to current theories of eye movements in reading. In J. Findlay, R. Kentridge, & R. Walker (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 381-392). Amsterdam, NL: NorthHolland.

Wade, N. (2010). Pioneers of eye movement research. *I-perception, 1,* 33-68

Wade, N. & Tatler, B. (2009). Did Javel measure eye movements during reading? *Journal of Eye Movement Research, 2,* 1-7.

Wade-Woolley, L. & Siegel, L. (1997). The spelling performance of ESL and native speakers of English as a function of reading skill. *Reading and Writing: An Interdisciplinary Journal, 9,* 387-406.

Wanzek, J., Vaughn, S., Wexler, J., Swanson, E., Edmonds, M. & Kim, A. (2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities, 39,* 528-543.

Westbrook, C. (2014). *Unlock. Reading & writing skills*. Cambridge, UK: Cambridge University Press.

White, S. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 205-223.

White, J. (1998). Getting the learner's attention: A typographical input enhancement study. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 85-113). Cambridge, UK: Cambridge University Press.

Williams, J. (2005). Learning without awareness. *Studies in Second Language Acquisition*. *27*, 269-304.

Williams, J. (2013). Attention, awareness and noticing in language processing and learning. In J. Bergsleithne, S. Frota & J. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 39-57). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Williams, R. & Morris, R., (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339.

Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition*. *35*, 323-352.

Wolf, M. (2008). *Proust and the Squid: The story and science of the reading brain.* New York, NY: HarperCollins.

Wolf, M. & Barzillai, M. (2009). The Importance of Deep Reading. *Educational Leadership*, *66*, 6,32-37.

Yamashita, J. (2013). Word recognition subcomponents and passage level reading in a foreign language. *Reading in a Foreign Language*, *25*, 52-71.

Yamashita, J. & Ichigawa, S. (2010). Examining reading fluency in a foreign language: Effects of text segmentation on L2 readers. *Reading in a Foreign Language*, *22*, 263–283.

Yeung, S, Siegel, L. & Chan, C. (2013). Effects of a phonological awareness program on English reading and spelling among Hong Kong Chinese ESL children**.** *Reading & Writing, 26,* 681-704.

Yu, C., Yurosky, A. & Xu, T. (2012). Visual data mining: An exploratory approach to analysing temporal patterns of eye movements. *Infancy*, *17*, 33-60.

Ziegler, J., Perry, C. & Zorzi, M (2014). Modelling reading development through phonological decoding and self-teaching: Implications for dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369* (1634).

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics, 36*, 277-284.

Zutell, J. & Rasinski, T. (1991). Training teachers to attend to their students' oral reading fluency. *Theory into Practice, 30*, 211–217.

Appendices

Appendix A. Informed Consent Form

Informed Consent Form

Date: Sept. 1, 2013
Study Name: An eye tracking study of Arabic L1 students and skilled English L1 readers while reading English texts.
Researcher: M. Joan Oakley
Sponsors: College of the North Atlantic-Qatar and Lancaster University
Purpose of Research: To compare the eye movements of Arabic L1 and English L1 readers while reading English texts.
What you will be asked to do in the study:
1) Attend a calibration session of about 5 minutes to see if you are a suitable candidate for the eye tracking experiment.
2) Complete a short personal information sheet.
3) Read two short passages in English and answer comprehension questions while the eye tracker is recording the way your eyes move. This will take no longer than 30 minutes.
Risks and Discomforts: I do not foresee any risks or discomfort from your participation in the research.
Benefits of the Research and Benefits to You:
This research will contribute to the literature on the difficulty that Arabic L1 speakers have when reading English texts.
For students: You will see exactly how your eyes move when you read in English. This will help you develop strategies to become a better reader.
For teachers: You will see exactly how your students process English texts at the word level. This will allow you to introduce pedagogical interventions which hopefully will help your students to become more proficient readers.
Voluntary Participation: Your participation in the study is completely voluntary and you may choose to stop participating at any time. Your decision not to volunteer will not influence the nature of the ongoing relationship you may have with the researcher or the nature of your relationship with the College of the North Atlantic-Qatar either now, or in the future.

Withdrawal from the Study: You can stop participating in the study at any time, for any reason, if you so decide. Your decision to stop participating, or to refuse to answer particular questions, will not affect your relationship with the researcher, the College of the North Atlantic-Qatar, or any other group associated with this project.

Confidentiality: Unless you choose otherwise, all information you supply during the research will be held in confidence and unless you specifically indicate your consent, your name will not appear in any report or publication of the research. Your data will be safely stored in a locked facility and only the researcher will have access to this information. Confidentiality will be provided to the fullest extent possible by law.

Questions about the Research? If you have questions about the research in general or about your role in the study, please feel free to contact Joan Oakley, mobile 66513818, joan.oakley@cna-qatar.edu.qa This research has been reviewed by the Institutional Review Board, College of the North Atlantic-Qatar and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines and the Supreme Council of Health guidelines for the State of Qatar. If you have any questions about this process, your rights as a participant in the study, or for copies of the results of this study, please contact Dr. Michael Long either by telephone at (974) 495-2236, or by e-mail (mike.long@cna-qatar.edu.qa)

Legal Rights and Signatures:

I _____ consent to participate in the eye tracking study conducted by Joan Oakley. I have understood the nature of this project and wish to participate. I am not waiving any of my legal rights by signing this form. My signature below indicates my consent.

Participant

Signature _____ Date _____

Principal Investigator

Signature _____ _____ Date_____

Appendix B. Example of the signs and sentences

Teachers and workers can park here.

Students cannot leave their cars here.

Appendix C. Example of a reading text

Halal Cosmetics

1.

For Muslim women who feel they are violating the teachings of Islam by using certain beauty products, Layla Mandi has the answer: Halal cosmetics. Her innovative beauty brand, OnePure, is making an immediate impact on the Middle-Eastern cosmetics scene with its products that promise effective results while keeping their 'halal' certification.

2.

Layla Mandi, a self-employed businesswoman, hopes to fill what she sees as a gap in the cosmetics industry for Muslim women in the region. Mandi, a Muslim convert, started her make-up career 15 years ago in her native Canada. As the years passed and she became more adept in her field, she began to look beyond the labels of her skincare products and realised that the ingredients of many creams and lotions did not fit with her idea of Islam.

She felt constant frustration at mainstream cosmetic ingredients that were considered 'haram' by her Islamic faith. She then began researching alternative ingredients that would have the same results while giving Muslins peace of mind. Determined to create a 'halal' product, Mandi brought together a dermatologist and a chemist and told them that the products must be free of alcohol and animal residue.

3.

"I developed OnePure because I wanted to care for my skin but wanted to observe my religious responsibilities. So, I researched and developed this very special halal skincare line," said Mandi. Believed to be the first halal cosmetics brand in the Middle East, OnePure products are certified in Malaysia by an Islamic

body that also certifies meats and other consumer goods for Muslims. So far, Mandi has been selling her products online, on Saudi Airlines, and from a small boutique in Burj Dubai.

4.

However, halal cosmetics are not a new idea. According to the Halal Journal, approximately $ US 150 million worth of halal products pass through the UAE every year. But these items are not readily available to consumers. At the Halal Expo 2015, Raees Ahmed, director, said there was an excellent opportunity for halal cosmetics to take advantage of the booming demand. A recent survey, in Malaysia, showed that 57.6% of Muslims in Singapore and 37.7 % in Indonesia, both emerging markets, were aware of halal cosmetics and would buy them if they were available. Mr. Ahmed added that halal products were also becoming popular with non-Muslim buyers. They are clean, wholesome and there are no impurities that go into the manufacturing process. Everyone, not just Muslims, likes the idea of that.

5.

Ahmad Azudin, senior manager at the International Halal Integrity Alliance (IHI) said “There is a growing demand for these products and an increasing awareness with consumers about animal contamination. Mr. Azudin and his team

are working on implementing an international halal standard for cosmetics by the end of next year. We are developing production standards for skin care, oral products, and fragrance in compliance with the Sharia board at IHI.

6.

Some in the make-up industry have expressed scepticism about halal cosmetics. “I feel it’s more about marketing,” said Noura Hamdi, marketing manager of Body Shop in Dubai. “We don’t use any animal products in our Skincare line and I know of other companies who don’t either.” Are Mandi and others filling a gap or creating a new market? It’s for the individual consumer to decide.

(Coh metrix 9.3)

Appendix D. Tracking activities

Tracking Exercise 1

The underlined word is the key word. Read the line of words and underline the key word each time you see it.

bit	bit	bat	bat	bit	bit	bat
set	sat	sit	set	sat	sat	set
woman	woman	women	woman	women	women	woman
poor	pour	pour	poor	pour	poor	poor
toil	toil	tool	toil	toil	tool	toil
Moza	Mouza	Mouza	Moza	Mouza	Moza	Mouza
Mona	Mona	Muna	Mona	Muna	Muna	Mona
Nasser	Nassir	Nassir	Nasser	Nasser	Nassir	Nassir
laid	load	load	laid	load	laid	laid
pleat	pleat	plait	pleat	plait	plait	pleat
more	mere	mere	more	mere	more	mere
scone	scone	scene	scone	scone	scene	scone
fitter	fatter	fatter	fitter	fatter	fitter	fitter
grew	grew	grow	grew	grow	grew	grow
mess	mass	miss	mess	miss	mass	mess
three	there	there	three	there	three	three
throw	throw	threw	throw	threw	threw	throw
bike	bake	bake	bike	bake	bake	bike
balk	balk	bulk	bulk	balk	balk	bulk
putted	patted	patted	patted	putted	patted	putted

Tracking Exercise 5

The underlined word is the key word. Read the line of words and underline the key word each time you see it.

trucker	tracker	tracker	trucker	tracker	tracker	trucker
wasp	wasp	wisp	wasp	wasp	wisp	wisp
raffle	ruffle	raffle	raffle	ruffle	raffle	raffle
shine	shone	shone	shine	shone	shone	shine
harbour	harbour	harbor	harbour	harbor	harbor	harbour
cave	cave	cove	cave	cove	cave	cove
inhabit	inhibit	inhabit	inhabit	inhibit	inhibit	inhabit
grand	grand	grind	grind	grand	grind	grind
fellow	follow	fallow	follow	fellow	fallow	fellow
spring	sprung	sprung	spring	sprung	spring	spring
pudding	pudding	pudding	padding	padding	pudding	padding
silver	salver	silver	silver	salver	silver	salver
person	person	person	parson	person	parson	parson
coach	couch	couch	couch	coach	coach	couch
perched	perched	parched	parched	parched	perched	parched
must	most	most	must	mist	most	mist
then	than	than	then	then	then	than
list	last	lost	lest	list	lust	list
dime	dime	dame	dome	dame	dime	dome
same	same	some	same	some	same	some
person	persona	persona	person	persona	person	person
remind	remand	remand	remind	remind	remind	remand
ramble	ramble	rumble	ramble	ramble	rumble	rumble
shave	shove	shave	shave	shove	shave	shove
forgive	forgave	forgive	forgive	forgave	forgive	forgave

Tracking Exercise 15

The underlined word is the key word. Read the line of words and underline the key word each time you see it.

pick the berries	pick the berries	pack the berries	pick the berries	pick the berries	pack the berries
it's by the stack	its' by the stack	it's by the stick	it's by the stick	it's by the stack	it's by the stick
she's a miss	she's a miss	she's a mess	she's a mess	she's a mess	she's a miss
they're noisy	they're noisy	they're nosy	they're noisy	they're noisy	they're noisy
the last word	the last ward	the last word	the last word	the last word	the last ward
a pretty bride	a pretty braid	a pretty braid	a pretty bride	a pretty braid	a pretty braid
a lovely person	a lovely person	a lovely person	a lovely parson	a lovely person	a lovely parson
they hop a lot	they hope a lot	they hope a lot	they hope a lot	they hop a lot	they hop a lot
he went abroad	he went aboard	he went aboard	he went abroad	he went aboard	he went aboard
a sore lip	a sore lip	a sore lap	a sore lap	a sore lip	a sore lap
a cheap bill	a cheap bull	a cheap bull	a cheap bill	a cheap bill	a cheap bull
a rounded hill	a rounded hill	a rounded hull	a rounded hill	a rounded hull	a rounded hill
I lost the pen	I lost the pen	I lost the pan	I lost the pen	I lost the pan	I lost the pen
two toy cars	tow toy cars	tow toy cars	two toy cars	two toy cars	tow toy cars
I command you	I command you	I commend you	I command you	I commend you	I command you
find the date	find the date	find the data	find the data	find the date	find the date
an awful course	an awful curse	an awful curse	an awful course	an awful course	an awful curse
he went to see	he went to see	he went to sea	he went to see	he went to see	he went to sea
a poor paper	a poor piper	a poor piper	a poor piper	a poor paper	a poor paper
expensive soap	expensive soap	expensive soup	expensive soap	expensive soup	expensive soup
read the letter	read the letter	read the letter	read the latter	read the letter	read the letter
a naval base	a novel base	a novel base	a naval base	a naval base	a naval base
an angry mule	an angry mule	an angry mule	an angry mole	an angry mole	an angry mole
unable to work	enable to work	enable to work	unable to work	enable to work	unable to work
tea time	tea time	tee time	tea time	tee time	tee time
lost in a cave	lost in a cove	lost in a cove	lost in a cave	lost in a cave	lost in a cove
more pudding	more padding	more padding	more padding	more padding	more pudding
in a rush	in a rash	in a rash	in a rush	in a rush	in a rash

Appendix E. Sample from NGSL List

Spelling and Rapid Word Recognition NGSL 1-105

1 the	36 one	71 because
2 be	37 my	72 come
3 of	38 know	73 could
4 and	39 there	74 use
5 to	40 which	75 work
6 a	41 can	76 then
7 in	42 get	77 now
8 have	43 her	78 also
9 it	44 would	79 than
10 you	45 think	80 him
11 for	46 like	81 into
12 not	47 more	82 only
13 that	48 their	83 want
14 on	49 your	84 look
15 with	50 when	85 these
16 do	51 what	86 its
17 as	52 make	87 new
18 he	53 time	88 give
19 we	54 who	89 first
20 this	55 see	90 way
21 at	56 up	91 thing
22 they	57 people	92 any
23 but	58 some	93 over
24 from	59 out	94 right
25 by	60 me	95 after
26 will	61 good	96 find
27 or	62 other	97 day
28 his	63 year	98 where
29 say	64 well	99 most
30 go	65 our	100 should
31 she	66 very	101 need
32 so	67 just	102 much
33 all	68 them	103 how
34 about	69 no	104 back
35 if	70 take	105 mean

Spelling and Rapid Word Recognition NGSL 526-630

526 range
527 percent
528 themselves
529 organization
530 vote
531 front
532 measure
533 trade
534 therefore
535 finally
536 raise
537 wear
538 industry
539 explain
540 relationship
541 quality
542 accord
543 outside
544 wish
545 death
546 project
547 land
548 sign
549 boy
550 news
551 risk
552 total
553 couple
554 national
555 list
556 opportunity
557 act
558 sport
559 road
560 kill
561 serve
562 education
563 picture
564 likely
565 standard
566 benefit
567 stage
568 performance
569 rest
570 certainly
571 culture
572 focus
573 itself
574 arrive
575 employee
576 upon
577 voice
578 due
579 technology
580 field
581 air
582 material
583 current
584 teach
585 financial
586 century
587 society
588 analysis
589 limit
590 evidence
591 reduce
592 listen
593 usually
594 lie
595 foot
596 single
597 common
598 space
599 realize
600 former
601 animal
602 instead
603 similar
604 thus
605 address
606 leader
607 complete
608 arm
609 function
610 chance
611 mention
612 factor
613 contact
614 response
615 demand
616 exist
617 accept
618 save
619 opinion
620 pick
621 wrong
622 apply
623 compare
624 suppose
625 choice
626 structure
627 fight
628 relate
629 firm
630 feature

Appendix F. Examples of Phonemic awareness activities

1. Phoneme identification/count using Elkonin boxes

Teacher pronounces a target word slowly, stretching it out by sound.Teacher asks the students to repeat the word. Teacher then draw "boxes" or squares on a piece of paper, chalkboard, or dry erase board with one box for each syllable or phoneme.

Teacher has the students count the number of phonemes in the word, not necessarily the number of letters. For example, *wish* has three phonemes and will use three boxes. /w/, /i/, /sh/

2 a. Phoneme deletion-Initial sound

Teacher: Say *cat.*

Student: *cat*

Teacher: Now say it without the /k/.

Student: *at*

2b.Phoneme deletion-Final sound

Teacher: Say *seat.*

Student: *seat*

Teacher: Now say it without the /t/.

Student: *sea*

2.c Phoneme deletion-First sound of a consonant blend

Teacher: Say *slip.*

Student: *slip*

Teacher: Now say it without the /s.

Student: *lip*

2.d Phoneme deletion- Embedded sound of a consonant blend

Teacher: Say *play*.

Student: *play*

Teacher: Now ssy it without the /l/.

Student: *pay*

3. Syllable identification

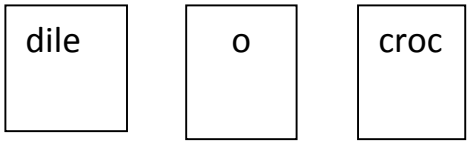
3 a. Separated syllables

Teacher writes words on the whiteboard syllable by syllable. Teacher asks students to use their knowledge of syllable spelling patterns (i.e closed syllables, open syllables, consonant +le etc.) to read each word.

ab sent	croc o dile	bot tle

3.b Multisyllabic word word manipulation

Teacher divides words from upcoming spelling list or reading into syllables. Teacher writes each syllabvel on a note card and displays the syllables in jumbled order. The teacher asks students to arrange the syllable to form a word.



3.c Deleting syllables

Teacher: The word is **lipstick**. Take off *stick*.

Student: *lip*

Teacher: The word is **butterfly**. Take off *fly*.

Student: *butter*

Teacher**:** The word is **racetrack.** Take off *race*.

Students: *track*

Teacher: The word is **sunshine**. Take off *sun*.

Student: *shine*

4. Rhyme

Teacher explains that rhyming words have final word parts with the same sound. Example:*house* and *mouse.* Teacher then reads several rhyme phrases or sentences aloud, emphasizing the word in bold and stopping before the rhyming word, and allow ssstudents to fill in the word.

1. The name of your **robe** is a _____. (thobe)

2. I hope that you **remember** your test is in _____. (September, November, December)

3. I do not want to **keep** my fourteen-year-old. _____ (Jeep)

4. He does not have **another**. He gave it to his _____. (brother,mother)

5. Our favourite sort of **mammal** must surely be the _____. (camel)

Appendix G Example of an enhanced reading text

Counterfeit Goods: Beyond Fashion

1.

Problems associated with counterfeiting are certainly not new. As far back as the Middle Ages, artisans used their own personal marks to distinguish their products from copied items. However, today's counterfeiters, have easy access to technological advances such as computers, copiers and scanners. Consequently, there are virtually no products that escape their reach. Never has it been so easy to duplicate labels, packaging, documentation and logos with such speed and accuracy. Counterfeiting has gone beyond Prada purses and Nike sneakers and now includes personal care products, fake prescription drugs, electrical items and auto and aviation parts.

2.

Bottles of fake perfume caused severe respiratory problems for consumers in the KSA. Laboratory tests showed that the products contained chemicals which can cause severe breathing difficulties when sprayed on the skin. Health authorities in Sierra Leone seized counterfeit toothpaste which contained high levels of a toxic chemical that could cause kidney and

liver failure. Fake cosmetics found in Dar es Salaam salons caused serious skin diseases which required medical attention for dozens of consumers. Even high-priced products can be cheap substitutes. Shoppers at a pharmacy in Jeddah bought an expensive skin care lotion and found themselves hospitalized for weeks. The phony lotion was found to contain toxic chemicals used in rat poison.

3.

Counterfeit pharmaceuticals are an area of even greater concern. They may contain dangerous ingredients or contain no active ingredients at all. According to the World Health Organization (WHO), 16% of all counterfeit drugs contain the wrong ingredients, 17% contain the incorrect amounts of the right ingredients and 60% of them have no active ingredients whatsoever. The effects of these products are being felt around the world. A fake blood thinning medicine killed 81 patients in the United States last year, while the same counterfeit medicine killed 76 in Australia. In 2006, WHO issued an alert to consumers warning that nearly 20,000 bottles of counterfeit Liptor (a drug used to control high cholesterol) had made their way onto the market. Last year, the chief of security for a major drug company testified in court that one counterfeit ring produced “millions of yellow tablets that were

virtually indistinguishable from the real product. The fake tablets were made of boric acid, floor wax and lead-based yellow paint used for road markings."

4.

Other counterfeit products constitute safety hazards. In 2007, a Canadian counterfeiter was jailed for supplying schools with unsafe lighting. Soon after they were installed, the lights began blowing out, showering the students with broken glass. In Thailand, an English tourist found his son dead on the floor of their hotel. The teenager had been electrocuted by a fake Nintendo charger that his father had bought in the market. In 2004, the US Consumer Safety Commission issued a recall alert warning consumers to immediately stop using counterfeit LG cell phone batteries. The batteries were susceptible to overcharging because they lacked a safety device. The recall notice warned that the fake batteries could overheat and pose a fire or burn hazard. In the UK, counterfeit lighters, with faulty flame -control mechanisms, caused severe burns to dozens of unsuspecting consumers. Counterfeit extension cords imported from China actually burst into flames when they were tested under normal household conditions.

5.

In the past 20 years, fake auto and aviation parts have caused countless deaths. Seven children died when the bus they were riding crashed into a brick wall. The brakes that had just been installed had a well-known trademark on them. However, on further inspection, they were found to be made of sawdust and soft plastic. An investigation into a Norwegian plane crash in 1989 found that fake substandard parts caused the plane's tail to fall off 22,000 metres above the North Sea. Even the US military has been fooled by counterfeiters. In 2004, Ralph Cooper pleaded guilty to selling **bogus** Black Hawk and Sea Hawk helicopter parts to the United States. Cooper bought parts from a Taiwanese company and relabeled them. The counterfeit parts that Cooper sold were made from substandard rubber that failed under the intense stress and heat experienced by military helicopters. Bogus parts have even been found on Air Force One (the plane used by US Presidents).

6.

Four people were killed and 54,000 others were poisoned in China when they drank contaminated milk which contained an industrial chemical. In Eastern Europe, counterfeit toys have caused brain damage in countless children due to the toxic paint used in their manufacture. Canada's RCMP found counterfeit cigarettes containing dead flies, insect eggs and mould. The list goes on. Every product in every industry is vulnerable to counterfeiting and piracy. Counterfeiting is a $250 billion industry that is putting consumers at risk. The criminals making fake goods care about making a profit; they don't care about consumer health and safety. The counterfeiter making a Louis Vuitton suitcase today could be making fake airplane parts tomorrow.

(Coh Metrix 4.10)

Appendix H - NGSL with meaning component

1. bury	to put something or someone in a hole in the ground
2. cure	to make an illness or problem better
3. decrease	to make something become less
4. melt	to turn from something solid to something soft or liquid
5. persuade	to make someone do or believe something by giving them a good reason
6. strengthen	to make stronger
7. border	the line that divides one country from another
8. courage	the ability to control your fear in a dangerous situation
9. passenger	a person who is travelling by bus, airplane, boat etc.
10. proof	information which shows that something is true
11. cotton	a material used to make clothing
12. lung	part of the body used for breathing
13. widow	a woman whose husband is dead
14. generous	giving a lot to other people
15. lonely	unhappy because you are not with other people
16. anxious	worried or nervous
17. stiff	hard or not easily bent
18. ugly	not beautiful

FL 1080 NGSL

Word	Word Form	Arabic	Meaning
1. bury			
2. cure			
3. decrease			
4. melt			
5. persuade			
6. strengthen			
7. border			
8. courage			
9. passenger			
10. proof			
11. cotton			
12. lung			
13. widow			
14. generous			
15. lonely			
16. anxious			
16. stiff			
18. ugly			

NGSL Vocabulary Matching

1. bury
2. cure
3. decrease
4. melt
5. persuade
6. strengthen

_____ to put something in a hole in the ground

_____ to make a person do something

_____ to make an illness or problem better

1. border
2. cotton
3. courage
4. passenger
5. proof
6. ton

_____ the line between two countries

_____ a person travelling by bus, airplane, boat, etc.

_____ Information that something is true

1. absent
2. horizon
3. lung
4. reflection
5. temple
6. widow

_____ a material used to make clothing

_____ part of the body used for breathing

_____ a woman whose husband is dead

1. ugly
2. anxious
3. generous
4. lonely
5. modest
6. stiff

_____ unhappy because you are not with others

_____ giving a lot to other people

_____ feeling nervous or afraid

Appendix I. Sample pages from the PHRASal ExpresssionsList (Martinez & Smidt,2012)

Appendix: The PHRASE List

Phrasal expressions divided to match 1K frequency bands of the most common word families in the BNC. The 'Integrated List Rank' represents where each item falls when both lists (individual and phrase lists) are merged together.

There are three categories of genre with frequency information to help the list user discern the appropriateness/usefulness of each phrase. The frequency information breaks down as follows:

* * * = phrase most common in this genre (or as common)

* * = phrase less common in this genre

* = phrase infrequent in this genre

X = phrase rare or non-existent in this genre

Integrated List Rank	**Phrase**	**Frequency** (per 100 million)	Spoken general	Written general	Written academic	**Example**
107	**HAVE TO**	83092	* * *	* *	*	I exercise because I **have to**.
165	**THERE IS/ARE**	59833	* * *	* * *	* *	**There are** some problems.
415	**SUCH AS**	30857	*	* * *	* * *	We have questions, **such as** how it happened.
463	**GOING TO (FUTURE)**	28259	* * *	* *	**X**	I'm **going to** think about it.
483	**OF COURSE**	26966	* * *	* *	*	He said he'd come **of course**.
489	**A FEW**	26451	* * *	* *	*	After **a few** drinks, she started to dance.
518	**AT LEAST**	25034	* * *	* *	* *	Well, you could email me **at least**.
551	**SUCH A(N)**	23894	* * *	* *	*	She had **such a** strange sense of humor.

5361	**TURN BACK**	827	*	***	*	There was no **turning back** now.
5366	**GET AWAY WITH**	824	***	***	x	Somehow they managed to **get away with** it.
5384	**NO WONDER**	820	***	**	x	**No wonder** he's ranked first.
5438	**WELL BEING**	809	***	***	*	It was good for their overall **well being.**
5447	**HOW ABOUT**	805	***	*	x	**How about** tomorrow?
5463	**TO GO ('REMAINING')**	801	***	**	x	Two more **to go.**
5479	**STRAIGHT AWAY**	794	***	***	*	They'll get to it **straight away.**
5485	**OWING TO**	793	**	***	***	**Owing to** his honesty, he was held in high esteem.
5491	**HOLD UP**	791	***	***	x	Sorry – we got **held up** on the M25.
5492	**LOOK TO**	790	**	***	***	He constantly **looked to** us for advice and support.
5496	**LAY OUT**	789	***	***	***	The countries **laid out** plans for a peace agreement.
5497	**THE LOT**	789	***	*	x	They tood the jewellery, the laptops, **the lot.**
5501	**KEEP ON**	788	***	**	x	Freeways **kept on** being built.
5502	**MAKE UP ONE'S MIND**	788	***	**	x	You'd better **make up your mind.**
5503	**AT WORK**	787	x	***	***	There were strange forces **at work.**
5504	**COME ABOUT**	787	**	***	*	It all **came about** through a meeting back in April.

----------**end of 5k**

TOTAL: 505