IELTS

IELTS Research Reports Online Series

ISSN 2201-2982
Reference: 2015/2

Using eye-tracking to research the cognitive processes of multinational readers during an IELTS reading test

Author: *Stephen Bax, Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire, UK*

Grant awarded: *2011–12*

Keywords: *"Eye-tracking research, IELTS reading test, cognitive processes of multinational readers, English language testing"*

Abstract

This article reports on a research project which used eye-tracking technology to investigate the eye movements of a group of multinational students completing IELTS (International English Language Testing System) test items. It represents the first attempt to use such technology to gain insights into the cognitive processes of students of different nationalities and languages as they read test passages and respond to test items.

The approach shadowed earlier successful research reported in Bax (2013a and 2013b). One limitation of that research was the use of a single nationality (Malaysian) group, leaving open the possibility that the cognitive operations of readers of other nationalities with different first languages, as revealed through eye movements and other methods, might be patterned in different ways. A further limitation of that research was that it was restricted to analysing local reading only. For this reason, the present study drew on the success of that earlier research, in terms of following its approach and methodology, but investigated a carefully selected multinational group and additional dimensions of their reading and test-taking behaviour not explored in the earlier study, through the use of the eye-tracking technology.

A cohort of multinational students (n=41) took an IELTS test which consisted of 11 test items and two authentic IELTS reading passages, delivered in onscreen mode to facilitate effective eye-tracking, carefully following the methodology of the Bax (2013b) study so as to allow for valid comparison. A random selection of these candidates was then made for eye-tracking analysis (n=30), and a sample of the same candidates (n=20) followed a retrospective stimulated recall procedure in which they reported on their reading. As in the earlier study, comparison was then made between successful and unsuccessful test candidates in terms of their eye movements and verbal reports.

The findings from this multinational group complement and extend the earlier research on a single nationality group in important ways. Significant differences were identified between successful and unsuccessful test-takers on a number of dimensions, differing in some respects from the findings of the earlier study. Areas of commonality included aspects of expeditious reading (Khalifa and Weir 2009), and various ways in which successful and unsuccessful readers focus differently on particular aspects of the test items and texts.

The research, therefore, offers significant additional insights from this new technology into the cognitive processing of multinational IELTS candidates in ways which could improve our development of reading test items, and also our preparation of candidates for reading tests.

Publishing details

Published by the IELTS Partners: British Council, Cambridge English Language Assessment and IDP: IELTS Australia
Web: www.ielts.org

AUTHOR BIODATA

Stephen Bax

Stephen Bax is Professor of Applied Linguistics at the Centre for Research in English Language and Assessment (CRELLA) at the University of Bedfordshire in the UK. He was awarded the 2014 TESOL Distinguished Researcher Award for a 2013 article in *Language Testing* which used eye-tracking to investigate reading tests, and his work also includes research into discourse, intertextuality and teacher education. His research into text analysis forms the basis for the *Text Inspector* online analysis tool, and he received an Elsevier prize for his work on Normalization in CALL. He has taught and researched in the Middle East, Asia and Africa.

IELTS Research Program

The IELTS partners, British Council, Cambridge English Language Assessment and IDP: IELTS Australia, have a longstanding commitment to remain at the forefront of developments in English language testing.

The steady evolution of IELTS is in parallel with advances in applied linguistics, language pedagogy, language assessment and technology. This ensures the ongoing validity, reliability, positive impact and practicality of the test. Adherence to these four qualities is supported by two streams of research: internal and external.

Internal research activities are managed by Cambridge English Language Assessment's Research and Validation unit. The Research and Validation unit brings together specialists in testing and assessment, statistical analysis and item-banking, applied linguistics, corpus linguistics, and language learning/pedagogy, and provides rigorous quality assurance for the IELTS test at every stage of development.

External research is conducted by independent researchers via the joint research program, funded by IDP: IELTS Australia and British Council, and supported by Cambridge English Language Assessment.

Call for research proposals

The annual call for research proposals is widely publicised in March, with applications due by 30 June each year. A Joint Research Committee, comprising representatives of the IELTS partners, agrees on research priorities and oversees the allocations of research grants for external research.

Reports are peer reviewed

IELTS Research Reports submitted by external researchers are peer reviewed prior to publication.

All IELTS Research Reports available online

This extensive body of research is available for download from www.ielts.org/researchers.

INTRODUCTION FROM IELTS

This study, by Stephen Bax of the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire, was conducted with support from the IELTS partners (British Council, Cambridge English Language Assessment and IDP: IELTS Australia) as part of the IELTS joint-funded research program. Research funded by the British Council and IDP: IELTS Australia under this program complements those conducted or commissioned by Cambridge English Language Assessment, and together inform the ongoing validation and improvement of IELTS.

A significant body of research has been produced since the joint-funded research program started in 1995, with over 100 empirical studies receiving grant funding. After undergoing a process of peer review and revision, many of the studies have been published in academic journals, in several IELTS-focused volumes in the *Studies in Language Testing* series (http://www.cambridgeenglish.org/silt), and in *IELTS Research Reports*. To date, 13 volumes of *IELTS Research Reports* have been produced. But as compiling reports into volumes takes time, individual research reports are now made available on the IELTS website as soon as they are ready.

Although eye-tracking studies have been used to investigate general reading processes for over a hundred years, it is only relatively recently that researchers have begun to use this methodology to investigate reading in a second or foreign language. Recent initiatives have started to explore the use of eye-tracking technology in the field of language testing with promising initial findings (e.g., Bax, 2013a, 2013b; Bax & Weir, 2012; Brunfaut and McCray; 2014; McCray, 2013; McCray, Brunfaut, and Alderson, 2012; Suvorov, 2015; Winke and Lim, 2014).

In the current study, Bax again attempts to investigate the differences in the cognitive processes of successful and less successful second language readers. However, in the current study he extends the research presented in Bax 2013a, 2013b, by addressing two perceived limitations of the original study; firstly, while the original participants were all of a single nationality (Malaysian), in the current study, a multinational group of participants was investigated. Secondly, the original study focused exclusively on 'local reading' but, as one of the major constructs of academic reading is considered to be 'global reading', the current study attempted to investigate this aspect of reading also.

However, as Bax notes, IELTS reading items which focus on global reading are not clustered in particular sections and, therefore, it was decided that it was not possible to research this feature directly. Instead, Bax hypothesised that differences might be found between successful and less successful readers at the pre-reading stage. Unfortunately he found that the learners' pre-reading activities were so diverse that it was impossible to distinguish between careful and expeditious reading in terms of eye movements. Consequently, he was unable to come to any satisfactory answer concerning global reading behaviour. Whether this is due to the design of the IELTS reading test itself, or is a limitation of using eye-tracking to reveal readers' cognitive operations, is no doubt something that could usefully be investigated in future research.

Dr Vivien Berry
Senior Researcher
English Language Assessment
British Council, London

References to the IELTS Introduction

Bax, S, (2013a) 'Readers' cognitive processes during IELTS reading tests: evidence from eye-tracking'. *British Council, ELT Research Papers,* pp 13–06, Available at: https://www.teachingenglish.org.uk/sites/teacheng/files/D069%20ELTRA%20report_FINAL_WEB%20ONLY.pdf

Bax, S, (2013b) 'The cognitive processing of candidates during reading tests: Evidence from eye-tracking', *Language Testing,* October 2013, vol. 30 no. 4, pp 441–465

Bax, S and Weir, CJ, (2012)'Investigating learners' cognitive processes during a computer-based CAE Reading test', *Cambridge Research Notes*, Cambridge ESOL, Issue 47, February 2012, pp 3–14, http://www.cambridgeesol.org/rs_notes/rs_nts47.pdf

Brunfaut, T and McCray, G, (2014). *Looking into reading: The use of eye-tracking to investigate test-takers' cognitive processing*. Paper presented at the Language Testing Forum 2014, University of Southampton, UK

McCray, G, Brunfaut, T and Alderson, JC (2012) *Combining eye-tracking with post test interview data to examine gap-fill items: triangulation of tribulation*, European Association for Language Testing and Assessment conference (EALTA), Innsbruck, Austria.

Suvorov, R, (forthcoming) *Interacting with visuals in L2 listening tests: An eye-tracking study.* ARAGS Research Reports Online. British Council, London

Winke, P and Lim, H, (2014) *The effects of testwiseness and test-taking anxiety on L2 listening test performance: A visual (eye-tracking) and attentional investigation.* IELTS Research Reports Online Series, IELTS partners, http://www.ielts.org/pdf/Winke%20and%20Lim.pdf

CONTENTS

List of graphics

List of tables

1 INTRODUCTION

Bax (2013b) reported on an innovative study which investigated the eye movements of a single nationality (Malaysian) group as they completed selected IELTS reading test items. That research (which was based in turn on a British Council ELTRA-funded study reported in Bax 2013a) aimed at gaining insights into the differences in cognitive processes between successful test-takers among the Malaysian cohort, and less successful ones, through the use of stimulated recall instruments in conjunction with extensive data from eye-tracking. Although that study was successful in identifying certain areas where successful readers differed from less successful readers in terms of cognitive processing, a significant limitation of the study was that it was restricted to one nationality group. It is known that eye movements of readers with different first languages (e.g. Arabic, Chinese) could potentially differ in important ways (Weger and Inhoff, 2006). For this reason, it is important to research IELTS reading through eye-tracking not only with a single nationality group, as in the earlier study, but with a multilingual group, and this was a principle aim of this study.

A second limitation of the earlier study was in its exclusive attention to local reading. As will be discussed below, a crucial aspect of the construct of academic reading is what is termed 'global reading', as opposed to 'local reading' only. Therefore, another important aim of this project was to investigate global reading, as well as local reading.

2 COGNITIVE VALIDITY IN READING TESTS

This study takes place in the context of what is termed 'cognitive validity', widely accepted as an essential requirement for language tests (Glaser 1991, Baxter and Glaser 1998). Field has recently defined this for listening as:

> *the extent to which the tasks employed succeed in eliciting from candidates a set of processes which resemble those employed by a proficient listener in a real-world listening event.*
> (Field 2013, p 77).

In terms of IELTS reading, the need to demonstrate cognitive validity likewise implies the need to research and ascertain the extent to which the IELTS reading tasks elicit reading processes from candidates which resemble those of readers in relevant real world contexts, such as in academic reading. To put it another way, in order to claim cognitive validity in its reading tests, IELTS needs to demonstrate that these tasks elicit cognitive processes which parallel, for example, the reading processes of university students in academic contexts. One anticipated contribution of this article, then, to build on the evidence set out in Bax (2013b) is to explore the potential contribution of eye-tracking to help assess the cognitive validity of IELTS reading test items, in this case with multilingual students.

3 MODELLING COGNITIVE PROCESSING IN READING

Previous research investigating cognitive processing in IELTS reading tasks includes Weir, Hawkey, Green and Devi (2009), which used retrospective questionnaires and reports to investigate the cognitive processes underlying the academic reading construct as measured by IELTS. That research built on the model of cognitive processing in reading tests proposed by Khalifa and Weir (2009), which in turn drew on work by Urquhart and Weir (1998) characterising reading as taking place at the *local* or *global* level, and being in nature either *careful* or *expeditious*. Global comprehension refers essentially to the understanding of information beyond the sentence, including the links between ideas in the text, and the various ways in which these are established. By contrast, local comprehension refers to "the understanding of propositions at the level of micro-structure" (Weir and Khalifa 2008, p 2).

Khalifa and Weir's model (2009) is important because it is the first substantive attempt to set out in a coherent way the relative levels of difficulty of cognitive processes in reading. This enables the investigation not only of whether a reading test covers all areas of cognitive processing, but also whether examinations aimed at advanced levels are appropriately addressing higher order cognitive operations as well as lower. For reference, the hierarchy proposed by Khalifa and Weir (2009) is reframed in Table 1, with additional glosses illustrating the ways in which the different levels can be operationalised. (Note: A 'gloss' is "a brief definition or synonym of unknown words provided in text in L1 or L2". Nation, 2002, pp 174)

Using questionnaire and *post hoc* recall to shed light on readers' cognitive processing can be effective, as Weir, Hawkey, Green and Devi (2009) demonstrated. However, within this process is the danger of participants potentially reporting their processing inaccurately, owing to the time lag between test-taking and their completion of the questionnaire. The use of immediate eye-tracking has obvious advantages both in its ability directly to record test-takers' eye activity as they read, and also in its recordings of participants' own eye movements which can then to be used as a reminder and stimulus in their *post hoc* recall of their own cognitive operations. Bax and Weir (2012) and Bax (2013b) found that eye-tracking was indeed effective in offering insights into candidates' cognitive processing, even though some aspects of that processing remained opaque. Therefore, it was decided in this project to follow the methodology and procedures adopted in the earlier (2013b) study, with the addition of new research questions, and using a multilingual participant group for reasons outlined above.

	Level of activity ***(ordered from more simple to more complex)***	**Readers' typical cognitive operations in language tests**	**Size of typical unit**
1	Lexis: Word matching	identifies same word in question and text	Word
2	Lexis: Synonym and word class matching	uses knowledge of word meaning or word class to identify a synonym, antonym or other related word	Word
3	Grammar/syntax	uses grammatical knowledge to disambiguate	Clause/sentence
4	Propositional meaning	uses knowledge of lexis and grammar to establish meaning of a sentence.	Sentence
5	Inference	goes beyond literal meaning to infer further significance	Sentence/ paragraph/ text
6	Building a mental model	uses several features of the text to build a larger mental model	Whole text
7	Understanding text function	uses genre knowledge to identify text structure and purpose	Whole text

Table 1: Levels of cognitive processing in reading tests (adapted from Khalifa and Weir 2009)

4 EYE-TRACKING IN READING RESEARCH

4.1 Eye-tracking and 'default' reading

For over a century, researchers have attempted to observe eye movements in reading (Wade 2010). A full review of eye-tracking research and its contribution is offered by Rayner (1998). Bax (2013b) has recently reviewed the work relating to eye-tracking in second language reading. It was noted in the latter review that no previous research had used eye-tracking to research second language reading under test conditions, and furthermore that the vast majority of eye-tracking research into reading has researched a very restricted form of reading termed the 'default mode' (Reichle et al. 2009), in which the reader's eyes proceed forward along the line of text with almost no regressions or difficulties, in a manner which is markedly different from second language reading under test conditions. In particular, test-takers frequently shift their gaze away from the reading text in order to look at and take account of the test items themselves, which again distinguishes this mode markedly from that of 'default mode' reading.

For this and other reasons, Bax argued that although there are certain elements of previous eye-tracking research which can inform eye-tracking research into reading under test conditions, the majority of findings derived from previous eye movement research cannot be directly applied to research into second language reading under test conditions. Despite this limitation, it is possible to isolate some general elements identified in previous eye-tracking research which Bax (2013b) suggested are still relevant to researching second language reading under test conditions.

These include the following aspects (cf. Rayner 1998, Rayner, Pollatsek, Ashby and Clifton 2012):

- *eye fixations* (when the eye dwells on a particular point) typically last about 200–250 ms, the mean saccade size (i.e. when the eye moves from one point to another) being 7–9 letter spaces (Rayner 1998, p 375)

- *saccades* (when the eye jumps from fixation to fixation). These can usefully be distinguished according to their function in reading, and can usefully be categorised into five types:

  - *rightward saccades* (in left-to-right languages), which take the eye onwards through the text
  - *regressions* i.e. backward movements through the text, usually aiming at correcting erroneous or ineffective processing (Rayner 1998)
  - *return sweeps*, which consist of the eye's return to a specific fixation point, perhaps one identified by the reader as the source location of a problem (associated with higher-proficiency readers)
  - *backtracking*, considered a strategy used by less effective readers, since it is less targeted than return sweeps, consisting of a more speculative movement back through a text, reminiscent of skimming as opposed to scanning
  - *corrective saccades*, defined as movements of the eye which successfully re-identify text (Rayner 1998); these are considered a mark of higher proficiency readers.

One interesting implication of this previous research into different types of saccades relates to their implied value in distinguishing more effective from less effective readers. More effective readers, it is suggested, will be more purposeful and focused in their reading (as evidenced by an apparently higher frequency of *return sweeps* and *corrective saccades*) while less effective readers will tend to be more speculative and unfocused (as evidenced by greater use of *backtracking*).

Although a close examination of saccade patterns is outside the scope of this article, it will nonetheless be possible to investigate in this project whether or not this broad characterisation of the behaviour of more successful as opposed to less successful readers can be supported.

4.2 Eye-tracking and cognitive processing

It has previously been argued that "eye movement data reflect moment-to-moment cognitive processes" (Rayner 1998, p 372), and other studies have supported the view that eye movements can help in investigating underlying mental operations (e.g., Bertram 2011, Buscher, Biedert, Heinesch and Dengel 2010, Eger, Ball, Stevens and Dodd 2007). Bax (2013b) reviewed this body of research and suggested that, although in broad terms it appears to be accepted that eye movements can assist in understanding mental processing, it is important to be cautious in assuming too direct a relationship between the two without further evidence from, for example, readers' own reports on their own processes while reading. Bax also noted that some recent computational models of eye movements, such as the latest versions of the E-Z Reader (Reichle et al. 2009, Rayner, Pollatsek, Ashby and Clifton 2012), caution against drawing large conclusions beyond the lexical coding level of reading, mainly owing to the limited evidence which eye-tracking can provide for higher order processes.

However, Bax also argued (citing studies such as Hyönä and Pollatsek, 1998 and Pollatsek, Hyönä and Bertram, 2000, in Reichle et al. 2009) that research into non-default or more 'disrupted' forms of reading, including the research into IELTS reading in the earlier research and also in this study, could potentially offer insights beyond the lexical level. One reason for this is that when researching reading during a language test, we have access to evidence not available to researchers of 'default mode' reading, in particular the participants' success or failure on the reading test items themselves. These provide a unique insight into whether or not, and to what extent, the readers understood key elements of the texts in question. If we also utilize other research tools, such as retrospective reporting, we can then legitimately infer whether a reader has used high-order inferencing strategies in his or her response to the test item. For this reason, the current project followed the methods utilized successfully in Bax (2013b) by distinguishing between readers on the basis of their test scores on each item, and also by including retrospective stimulated recall as a central part of its methodology.

5 RESEARCH METHODOLOGY

5.1 Research questions

For reasons outlined above, this study aims to use eye-tracking technology to go beyond the earlier research in Bax (2013b) in two ways, firstly by researching a mixed nationality/language group and secondly to look at global reading as well as local reading. In other words, whereas the research described in Bax (2013b) examined only 'careful local' and 'expeditious local' comprehension, it was decided in this study also to attempt to research aspects of global reading behaviour, as evidenced in eye movements.

It was appreciated from the outset that since this area has not previously been investigated using eye-tracking, this direction was both innovative and potentially difficult, partly since the terms 'careful' and 'expeditious' have not yet been defined in terms of trackable eye movements. Nonetheless, in view of the importance of this aspect of reading in academic study, it was decided to investigate the following two questions relating to global reading, in addition to researching the areas of local reading described below:

Research question 1: Do successful students read the whole text carefully (careful global reading) more than unsuccessful students?

Research question 2: Do successful students skim the whole text speedily (expeditious global reading), more than unsuccessful students?

A further three research questions were adapted from the earlier study (Bax 2013b), as follows:

Research question 3: To what extent and in what ways can eye-tracking technology shed light on cognitive processing of participants completing IELTS Academic reading test items onscreen?

Research question 4: To what extent and in what ways are successful readers differentiated from less successful readers in terms of their eye movements while completing IELTS Academic reading test items onscreen?

Research question 5: To what extent and in what ways are successful readers differentiated from less successful readers in terms of their cognitive processing while completing onscreen reading test (IELTS) items, as evidenced from eye movement data and stimulated retrospective interview data?

5.2 Research approach and instruments

In view of the success of the earlier study (Bax 2013b), and in order to compare the two studies, it was decided to follow the main approach of that project.

5.3 Participants

An onscreen version of two IELTS reading passages with a total of 11 test items, selected so as to target the relevant cognitive processes, was delivered to a cohort of international students (n=41), from Africa, East Asia, Central and Eastern Europe and the Arab world, with 21 different first languages. The layout of the items on screen can be seen in Appendix 1. Participants' scores ranged from IELTS Band 5.5 – Band 7.0 with an average score of Band 6.0, and were drawn from Foundation Year and first-year undergraduates studying at a UK university. A random group of participants (n=30) were selected for eye-tracking in ways described below, and all their activities were captured using screen recording software.

Text topic	Number of test items	Type of items	Cognitive process targeted cf. Khalifa and Weir (2009)
Human Genome Project	5	Sentence completion- select words from the passage – constructed response.	Careful local reading
Biometric Security Systems	6	Matching	Expeditious local reading

Table 2: Characteristics of the selected reading test texts and items (Devi 2010)

5.4 Preliminaries

After completing ethics procedures and information forms, the participants completed computer familiarity questionnaires during which they all reported extensive familiarity with computer technology and onscreen tests of various kinds. For this reason, it was determined that the test mode had no significant impact on test-takers' behaviour.

5.5 Texts and test items

Research questions 1 and 2: Global reading

As discussed above, this project aimed to investigate global reading using eye-tracking in a way not previously attempted, as well as local reading. Since the IELTS reading test does not have a particular section or set of questions devoted to testing global comprehension, it was not possible to research this dimension directly. However, previous research indicates that successful IELTS students frequently read the texts through *before* attempting the test items, with a view to grasping global meaning, as is clear from this account by an IELTS candidate.

> "I usually read the texts carefully from the beginning to the end initially then I go to the questions. I can answer some questions without having to read the text again. If not, I usually remember the place where the info necessary for the answer is located and go there usually by scanning which may be followed by some careful reading." (Cited in Weir, C, Hawkey, R, Green, A, Unaldi, A and Devi, S, 2012, p 86)

On the assumption that this might represent a pattern of behaviour among IELTS students more generally, it was decided to analyse the eye movements of those students who read the text before looking at the test items, defining this type of reading as 'global' since it aims at obtaining a global representation of the text as preparation for later attempts to answer each question. It was anticipated that detailed examination of this segment of the eye-tracking data might allow a principled distinction to be drawn between 'careful global reading' and 'expeditious global reading' (e.g. 'skimming') on the basis of differences in total reading time, total number of fixations, saccade lengths or similar. Any patterns which could be identified along these lines could then potentially be correlated with students' wider success or failure on answering the text items themselves, so as to answer the first two research questions posed above.

Research questions 3, 4 and 5: Local reading

Investigation of the remaining three research questions was carefully designed to follow the procedures described in Bax (2013b) to allow for valid comparison, including the use of the same test items. As noted in the earlier paper, the texts and task had previously been piloted by Devi (2010) and were selected from the Academic version of IELTS Practice Papers series (Cambridge University Press), having been developed and trialled by Cambridge ESOL (English for Speakers of Other Languages). The test items chosen are described in Table 2, and the precise items used are listed in Appendix 2 below.

5.6 Test delivery

For more accurate tracking of participants' eye movement, the IELTS texts and items were transferred into onscreen format. Participants had already indicated familiarity with onscreen test-taking but so as to ensure that they were fully comfortable with the delivery mode, a detailed pre-test training video on the format and delivery of the test was presented to each test-taker.

5.7 Eye-tracking technology specifications

In terms of the eye-tracking equipment, the same device as in Bax (2013b) was used so as to ensure full comparability across the two studies. This was a Tobii T60, which has a sample rate of 60 Hz per second, set to a screen recording rate of 10 frames per second. (Full technical specifications can be found at: http://www.tobii.com)

5.8 Procedure

Again, so as to allow valid comparison with the earlier single nationality study, a similar procedure was adopted with this multinational group. After completion of all personal information forms, consent forms and computer familiarity forms, the project followed the following steps.

Stage 1: Individual eye movements were calibrated for each participant using the Tobii calibration tool. This ensured the accuracy of the device's tracking of their reading during the test.

Stage 2: Each participant watched a short video tutorial, with step-by-step explanation of the process they were about to follow.

Stage 3: Each participant then completed the IELTS reading items onscreen.

Stage 4: A random sample of participants (n=20) then completed a Retrospective Stimulated Recall interview procedure, described below (Section 5.9).

5.9 Stimulated recall interviews

Given the danger noted above of assuming too direct a relationship between eye movements and cognitive processes, it was important in addition to the eye-tracking record itself to obtain participants' reports on the processes they had followed while reading. For this reason, a random sample of the eye-tracked candidates (n=20) underwent a Stimulated Recall Interview. One benefit of eye-tracking technology is that it provides a visual record of the second-by-second eye activity in video format, and this was used as a stimulus to participants who then explained and commented on their reading activity at each point. These reports constituted important evidence on readers' cognitive processes, and served to elucidate and amplify the eye-tracking data.

6 ANALYSIS AND FINDINGS

6.1 Item analysis

In order to check the reliability of the items, item analysis was conducted, with results summarized in Table 3. It will be seen that the reliability coefficient was .798 (Cronbach's Alpha), which is generally considered acceptable considering the limited number of items under scrutiny (Pallant 2010). On this evidence, the test items appear relatively easy for the test population, although the means, (with the most difficult items 1 and 5, showing a mean of .41, and the easiest item, 9, as .80) suggest that they were nonetheless still targeting the participating students' proficiency levels reasonably well.

6.2 Analysis of the eye-tracking data

6.2.1 Quantitative analysis

The eye-tracking data consisted of full recordings of 30 participants' complete eye movements. The process of analysis, aligned with the earlier research on the single nationality group (Bax 2013b), proceeded as follows.

Step 1: Each test item was carefully analysed so as to determine the cognitive processes which a reader would need to employ in order to answer the item correctly. The full analysis can be seen in Appendix 3. Some items in the test (e.g. item 1) required cognitive processing at relatively lower, lexical levels. By contrast others (e.g. item 5) also required additional higher-order cognitive processes such as inferencing (see Level 5 of Table 1).

	Mean	Std. Deviation	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
item01	.41	.499	.494	.778
item02	.68	.471	.379	.790
item03	.46	.505	.351	.794
item04	.49	.506	.634	.762
item05	.41	.499	.416	.787
item06	.68	.471	.419	.786
item07	.71	.461	.345	.793
item08	.76	.435	.562	.772
item09	.80	.401	.469	.782
item10	.54	.505	.458	.782
item11	.51	.506	.504	.777

Table 3: Item analysis of the 11 Reading items (N=41)

Step 2: Following this analysis, it was possible to identify and locate Areas of Interest (AOIs) in the texts themselves, namely words or phrases which readers would have to locate and use in order to answer each test item. A suitable margin of error was allowed in terms of the space around the word on which the eye might fixate, so as to account for individual variation. Identification of AOIs then allowed the software to calculate the relative quantity of fixations used by successful and less successful test-takers, to see whether more successful readers looked more frequently at particular key words than less successful readers, as well as other important eye movement behaviour.

Step 3: The eye-tracking software then generated statistical data allowing for detailed comparison of test-takers' behaviour in order to investigate Research questions 3, 4 and 5 (Section 5.1 above). Eye movements of those participants who were correct on that item were compared with eye movements of participants who were incorrect on that item. This allowed analysis of whether differences in participants' eye movement might contribute to their failure or success in each test item. As in Bax (2013b), this included the calculation and comparison of the attention paid by each reader to these areas.

For Research questions 1 and 2, relating to global reading, the text as a whole was treated as the unit of analysis in the case of those participants who read it before reading the test items, the aim being to see if differences could be found in the global reading processes of successful and unsuccessful participants.

For Research questions 3 to 5, relating to local reading, the following areas were investigated.

a) The text as a whole, on the understanding that this might provide insights into each reader's ability to read expeditiously, with more successful readers able more effectively to find the crucial parts of the text in their search for an answer.

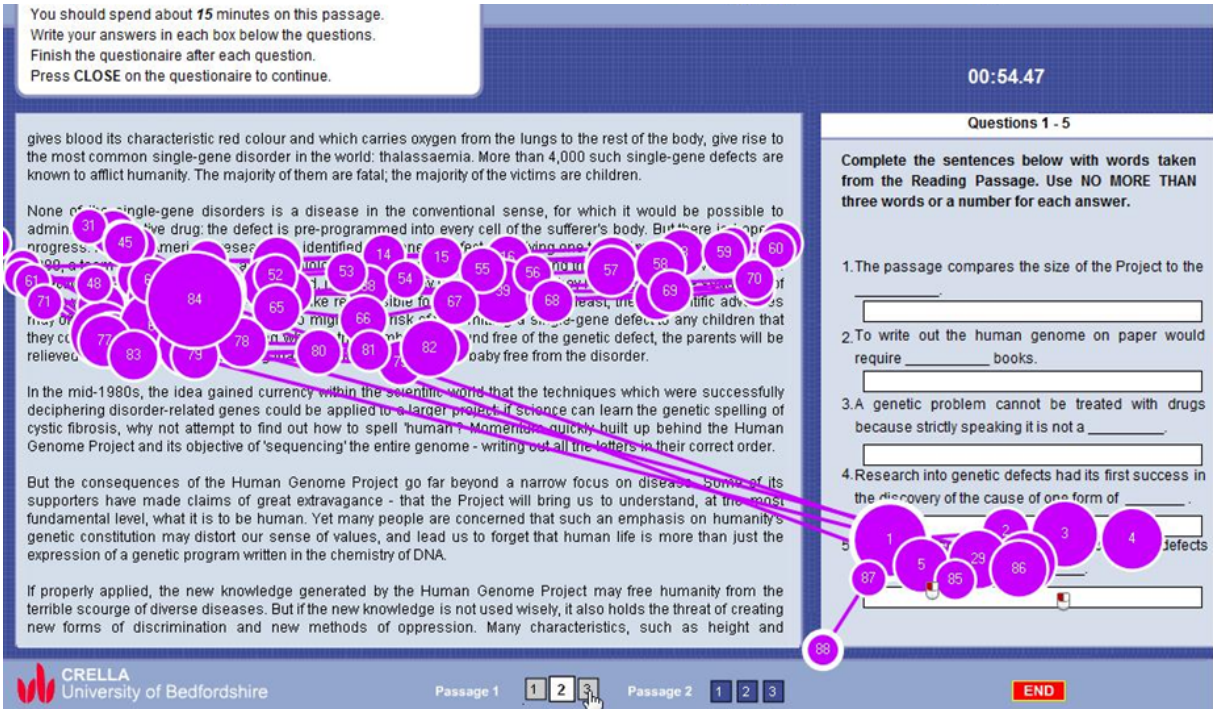
b) Key sections of the text at sentence level or beyond; if successful readers spent significantly longer on these, this might be attributed to stronger ability at relatively higher cognitive levels.

c) Specific areas of the text and test item previously identified as targets (Areas of Interest or AOIs); relative differences between successful and unsuccessful readers in this sphere could possibly be attributed to the relative importance each gave to key lexis or grammar. Successful candidates might be expected to fixate more and for longer time periods on such key elements than less successful readers.

In each case, it was clear that the eye movement data would need corroboration from the *post hoc* retrospective recall data.

In order to compare the behaviour of successful and less successful test-takers on each item, the non-parametric independent samples Mann-Whitney U test was used, as in the previous study, since the datasets did not meet assumptions of normality and homogeneity of variances.

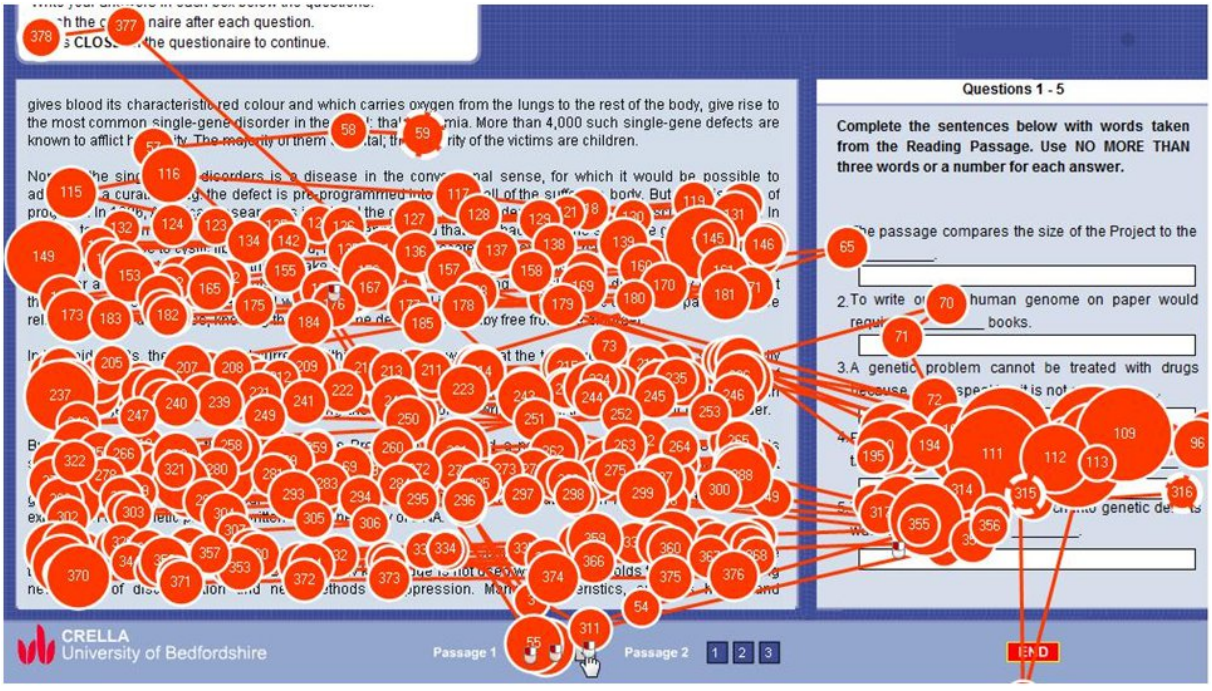
Step 4:
The eye-tracking software furnished tools which allowed for the detailed analysis of individual participants' reading behaviour. These included GazePlot data (as illustrated in Graphics 1 and 3) and Heatplot data (as illustrated in Graphics 2 and 4). Step 4, therefore, consisted of careful analysis of individual patterns, using such software tools in conjunction with the retrospective Stimulated Recall Interview data provided by unsuccessful and successful participants, so as to build a comprehensive picture of reader behaviour in those items where statistical significance was indicated. By way of example, it was apparent in the case of item 5 (as illustrated in the graphics below) that unsuccessful students spent significantly more time reading the text as a whole than did successful students. On the basis of the visual tools and interview data, it was possible to determine that the probable reason for this was that unsuccessful students, according to their self-reports, had found the relevant part of the text only with some difficulty, which implies weak abilities in expeditious reading skills. This incidentally confirmed similar findings in Bax (2013b) and was borne out in analysis of item 10, as will be seen below.



Graphic 1: Gazeplot output from successful candidate answering item 5

Graphic 2: Heatmap output from successful candidate answering item 5

In contrast with the graphics above, Graphics 3 and 4 show an unsuccessful candidate completing the same item, failing expeditiously to locate the place of the answer, and, therefore, spending much time wastefully scanning the page, taking more than 172 seconds, and answering incorrectly.



Graphic 3: Gazeplot output from unsuccessful candidate answering item 5

Graphic 4: Heatmap output from unsuccessful candidate answering item 5

7 FINDINGS

7.1 Global reading

It will be recalled that this study aimed to investigate not only careful and expeditious 'local reading' (for which the findings will be discussed below) but also expeditious and careful 'global reading' in the form of Research questions 1 and 2. Since it was not possible to research this directly, given that the IELTS reading items which focus on global reading are not grouped in particular sections, it was decided instead to investigate the eye movements of those participants who read the passage before reading the test items, on the expectation that they did so in order to obtain a 'global' sense of the whole passage. It was anticipated that a difference might be found in this pre-reading stage between those readers who were then more successful in the test as a whole and those who were not. If such a relationship could be found, it could be potentially of great interest to teachers, learners and test-takers.

However, results in this area were inconclusive. Of the students whose eye-tracking data were examined, it turned out that their reading activity prior to their reading of the test items was remarkably varied. In practice, with regard to both successful and unsuccessful students (in terms of their eventual test scores), only a handful read the whole text in advance, while some read part of it and others none at all. As a result, there were insufficient student samples to allow for comparable data in terms of global reading.

Furthermore, of those few who did read large parts of the text in advance, it was impossible in practice to distinguish in any principled way between careful and expeditious reading in terms of eye movements. While this might be possible in future with a larger sample, with this cohort there were not sufficient students who read enough of the text to allow for this central distinction to be drawn definitively. As a result, it was not possible to come to any satisfactory answer concerning global reading behaviour. Therefore, it was not possible to answer Research questions 1 and 2 in any satisfactory way.

7.2 Local reading

Research questions 3 to 5 concerned 'local reading'. The decision to shadow the earlier study (Bax 2013b) closely in terms of method and approaches allowed for the identification of important differences between the two, as well as some common findings. For example, in the single nationality study, significant differences were found between the eye movement activity of successful and unsuccessful participants in test items 2, 3, 5, 7 and 10. However, in the present study, using the same test items, no significant differences were found with regard to items 2, 3 and 7; by contrast significant differences were found in items 4 and 11. Both studies found significant differences in terms of items 5 and 10. In terms of item 5, discussed in detail below, the two research studies both identified the same significant difference in the case of one AOI, and in addition the present study identified a second area as well.

In the case of item 10, the present study identified an additional area of significant difference which was not present in the earlier study, in ways to be outlined below.

The implication of these interesting differences will be explored in the next sections, following presentation of the findings concerning those items where significant differences were found (items 4, 5, 10 and 11).

7.2.1 Item 4

Item 4 was as follows:

> *"Research into genetic defects had its first success in the discovery of one form of ____"*

The correct answer, 'muscular dystrophy', was located on the second page of the text in paragraph two. By that stage, most students had completed the first three questions, some of whose answers were on page 1 of the text, which meant that in order to answer item 4 correctly, readers had first to search expeditiously to find the location of the answer. They then needed to use lexical knowledge (synonymy) to match the phrase *'one form of'* in the question with '*one type of*' in the text. In addition, they needed to make use of higher order inferencing skills to interpret the dates in the text to work out which discovery was the 'first success' mentioned in the question. In short, they needed a moderately complex combination of cognitive processes, across the range of those identified by Khalifa and Weir (see Table 1 above).

The statistical data showed that candidates who were successful on this item focused significantly more, in terms both of fixation count and also of visit duration, on the synonymous phrases *one form of* and *one type of* in the question and text respectively (see Table 4), suggesting their higher attention to these important lexical elements. For example, it can be seen that the median fixation count for the phrase *one form of* in the question was 12.00 seconds for the successful students but only 3.00 seconds for unsuccessful candidates (Col 1, Table 4). For the corresponding phrase in the text (*one type of*), the median values were 3.00 seconds and 0 seconds respectively, showing that the majority of unsuccessful students did not focus on it at all. The measure for 'total visit duration' for these elements points in the same direction (Table 4). Successful students also focused more in term of fixation count (median of 17.00 as opposed to 2.00), and in terms of total visit duration (7.94 versus 0.47 seconds) on the key date of 1986, which was necessary to identify which was the first success of the research in question.

In summary, the eye-tracking data suggest that success with this item was due in part to a better ability to identify key lexical elements in the question and the text, and to identify and focus on key textual elements necessary to disambiguate and select an answer from a range of possible options. As noted above, the cognitive operations in question were those related to lexis (synonymy) and, at a higher level, to inferencing.

7.2.2 Item 5

Statistically, item 5 was the most difficult item, along with item 1 (see Table 3 above). But whereas the difficulty in item 1 came from the presence of a strong distracter in the text which confused many readers, the difficulty in item 5 derived from the requirement to employ cognitive operations at both the lexical (synonymy) level and also at a higher level of cognitive processing in terms of inference and reading across sentences.

The first step towards answering the question was to identify the location of the answer in the text. The eye-tracking data showed that on one measure, 'fixation count', unsuccessful readers spent significantly longer looking over the whole text than successful students, implying that they could not find the location of the answer efficiently, i.e. that they had weaker expeditious reading skills. Interview data confirmed that unsuccessful students had great difficulty in using expeditious reading strategies to identify where to find the answer in the text. This was also apparent in the visual data; as can be seen in the graphical examples above (Graphics 1-4), successful students located the answer efficiently, while unsuccessful students tended to read with less focus and purpose.

Measure	***Fixation count***			***Total visit duration***		
Target	Question element: *one form of*	Text element: *one type of*	Text element: *in 1986*	Question element: *one form of*	Text element: *one type of*	Text element: *in 1986*
Mann-Whitney U	35.00	38.50	37.00	47.50	36.00	32.00
Z	-3.232	-3.128	-3.154	-2.696	-3.193	-3.385
Sig. (2-tailed)	.001	.002	.002	.007	.001	.001
Mean (incorrect group n=15)	4.60	1.73	6.13	3.87	0.57	2.40
Median (incorrect group, n=15)	3.00	0.00	2.00	3.01	0.00	0.47
SD (incorrect group)	3.58	3.24	8.96	3.24	1.11	3.81
Mean (correct group n=15)	14.07	5.13	19.07	8.66	2.01	8.02
Median (correct group n=15)	12.00	3.00	17.00	7.44	1.52	7.94
SD (correct group)	11.93	4.76	13.47	5.19	1.53	5.48

(All significant at $p < 0.05$)

Table 4: Eye-tracking statistics for Item 4

Table 5 shows that the median fixation count for unsuccessful students was 367 seconds, but only 159 seconds for successful students. Again this supported the possibility that unsuccessful students were unable to locate the answer efficiently. However, on other measures of time allocated to the whole text, such as total visit duration, there was no significant difference found between successful and unsuccessful students.

		Fixation count Whole text
Incorrect (N=17)	Mean	352.47
	Median	367.00
	Std. Deviation	259.49
Correct (N=13)	Mean	202.31
	Median	159.00
	Std. Deviation	146.45
Mann-Whitney U		63.00
Z		-1.988
Sig. (2-tailed)		.047

(Significant at $p < 0.05$)

Table 5: Eye-tracking statistics for Item 5

A further significant difference here between successful and unsuccessful students, as with item 4 discussed above, was in the focus on key elements of the text, for which details can be seen in Table 6. It is clear from this that successful students did focus significantly more on key parts of the text essential to identifying the answer. Item 5 required readers to identify "*the cause of* one form of a disorder" (*cystic fibrosis*, the correct response to the item). This required them first to identify and comprehend the relevant section of the text, namely the phrase *gives rise to*. They also then had to locate and focus on the date *1989,* since that informed them that this was the second major discovery, and therefore the correct answer.

It is apparent from the evidence in Table 6 that successful students did in fact focus significantly more on these two elements in terms of fixation count and total visit duration, suggesting that they identified them successfully, then used them appropriately in proceeding towards the correct answer. By contrast, unsuccessful students did not focus adequately on those essential elements. In fact, it is revealing that more than half did not find them or focus on them at all, which is why the median scores are zero in each case.

The retrospective interview evidence supported this analysis. For instance a successful student reported that "I looked for the date because I wanted to see which was the first discovery and the second". Another said: "I looked at those words [*give rise to*] because that gave the meaning of the answer". In this way, eye-tracking and interview data both suggest that successful students correctly processed the relevant syntax and other crucial information in the text, whereas the unsuccessful students failed to do so.

7.2.3 Item 10

Findings on Item 10 coincide in an interesting way with those of the earlier study, in that here too the data showed successful students focusing significantly more on the precise paragraph in the text where the correct answer was located (Table 7). This supported the view that a key difference between successful and unsuccessful readers revolves around the inability of the latter to read expeditiously, particularly to locate the correct source of the answer. This is apparent in Table 7, which shows that successful students spent 16.95 seconds on the paragraph (total visit duration, median) in contrast with only 1.7 seconds for unsuccessful students. The same applies to fixation count, with successful students fixating 51 times (fixation count, median), or 55.77 times (mean) on the paragraph, and unsuccessful students fixating only 5 times (fixation count, median) or 13.62 times (mean).

		Fixation count on target in text*: gives rise to*	**Total visit duration on target*: gives rise to***	**Fixation count on target in text*: 1989***	**Total visit duration on target*: 1989***
Incorrect (N=17)	Mean	4.94	2.07	0.71	0.21
	Median	0.00	0.00	0.00	0.00
	Std. Deviation	8.61	4.29	1.10	0.33
Correct (N=13)	Mean	16.46	7.80	3.00	1.34
	Median	16.00	6.81	3.00	1.70
	Std. Deviation	9.05	4.96	2.48	1.25
Mann-Whitney U		30.00	28.00	51.00	52.50
Z		-3.434	-3.518	-2.668	-2.595
Sig. (2-tailed)		.001	.000	.008	.009

(All significant at $p < 0.05$)

Table 6: Eye-tracking statistics for Item 5

		Fixation count on correct paragraph	Total visit duration on correct paragraph	Fixation count on target word: *housing*	Total visit duration on target word: *housing*
Incorrect (N=13)	Mean	13.62	4.02	0.38	0.09
	Median	5.00	1.70	0.00	0.00
	Std. Deviation	22.26	6.51	0.87	0.20
Correct (N=17)	Mean	55.77	21.21	3.12	1.19
	Median	51.00	16.95	2.00	0.85
	Std. Deviation	39.41	17.14	3.66	1.43
Mann-Whitney U		37.50	35.00	55.00	54.00
Z		-3.07	-3.17	-2.53	-2.57
Sig. (2-tailed)		.002	.002	.012	.010

(All significant at $p < 0.05$)

Table 7: Eye-tracking statistics for Item 10

In addition, with item 10 in this study there was a significantly greater focus by successful participants on a key word in the text which was required for the answer. The question used the phrase *home owners* so that in order to locate the answer, participants had to find and interpret the word *housing* in the text and identify it as lexically related to *home*. Table 7 demonstrates that successful participants operated more effectively at this (lexical) level than unsuccessful readers. In fact, most unsuccessful participants on this item did not fixate at all on the word *housing*, which explains the median score of zero in Table 7. In other words, the eye-tracking data on item 10 gives evidence of successful participants' better expeditious reading, and also of better attention to lexical information in answering the question.

7.2.4 Item 11

The eye-tracking evidence for item 11 was similar to that for item 10, in that it showed successful participants to be more effective both at using expeditious reading skills, and also at focusing on key lexical elements needed for the correct answer.

In terms of expeditious reading, unsuccessful participants on this item spent significantly longer scouring the text as a whole, as measured by total fixation duration (see Table 8). This suggests that they were unable to correctly locate the answer in the text. Table 8 shows that they spent a mean of 108.21 seconds focusing on various aspects of the whole text, significantly more than successful students. At interview, several unsuccessful students reported that they had difficulty finding the location of the correct answer for this item, and spent more time looking through the text to find it. However, on other measures, such as total visit duration and fixation count, there were no significant differences.

By contrast, successful participants focused significantly more on the key word in the text (*customers*) that matched the same word in the question, which is at the lowest level of Khalifa and Weir's hierarchy (Table 1 above), i.e. lexical matching. Table 9 shows that successful participants focused an average of 2.27 times on the target word, as against 0.87 for unsuccessful participants. In fact, most of the latter group did not fixate on it at all, which explains the median score of zero. Clearly, if they fixated on this target word almost not at all, it explains in large part why they failed to find the correct answer.

		Total fixation duration on whole text
Incorrect (N=15)	Mean	108.21
	Median	78.67
	Std. Deviation	96.39
Correct (N=15)	Mean	48.00
	Median	34.14
	Std. Deviation	42.42
Mann-Whitney U		35.00
Z		-3.172
Sig. (2-tailed)		.002

(All significant at $p < 0.05$)

Table 8: Eye-tracking statistics for Item 11

		Fixation count on target element in text: *customers*
Incorrect (N=15)	Mean	0.87
	Median	0.00
	Std. Deviation	1.69
Correct (N=15)	Mean	2.27
	Median	1.00
	Std. Deviation	2.22
Mann-Whitney U		62.00
Z		-2.25
Sig. (2-tailed)		.025

Table 9: Eye-tracking statistics for target word in Item 11

8 DISCUSSION

The results from this study offer illuminating evidence concerning successful and unsuccessful test-takers' behaviour on these IELTS items, some of which confirms the findings from the earlier study, and some of which adds new insights. The findings are summarized for reference in Table 10.

In methodological terms, these findings confirm those of earlier studies (Bax and Weir 2012, Bax 2013b) in demonstrating the value of eye-tracking technology, in conjunction with other methods, including retrospective recall, in allowing the identification of differences in cognitive operations between successful and unsuccessful readers at a number of different levels of operation, in Khalifa and Weir's terms (2009, see Table 1).

In terms of cognitive validity, these IELTS test items succeeded in distinguishing between the cognitive processing of successful and unsuccessful student test-takers at the levels of lexis (word matching and synonymy) and syntactic processing. In other words, successful candidates did make use of the types of cognitive strategies which would be expected in real-life academic situations, whereas unsuccessful candidates did not, matching the findings of the earlier study. Of particular interest in terms of higher order processing was the fact that the two items which required readers to locate specific dates in order to build correct propositional information and make correct inferences (i.e. items 4 and 5) both succeeded in eliciting the correct processing from successful candidates, according to the eye-tracking evidence. The implication is that those participants who were unsuccessful on these two items failed in part because they did not locate the dates and so failed to draw the correct inferences.

In respect of the cognitive validity of these 11 IELTS test items as a whole, the result is even more positive if we consider the findings from the two studies together, since the number of items shown to elicit significant differences between successful and unsuccessful test-takers across the two studies is 7 of the 11 test items (items 2, 3, 5, 7, 10 and 11 in the first study and items 4, 5, 10, 11 in this study). Given that eye-tracking has recognized limitations in terms of the cognitive operations into which it can give insight (as discussed further below), this is an impressively high proportion.

8.1 Research questions 3-5

In terms of Research questions 3 to 5, the findings can be summarized as follows:

- Research question 3: Eye-tracking technology can shed light on cognitive processing of participants completing IELTS Academic reading test items onscreen to some extent, although for many items no significant differences were found in the eye-tracking data, for reasons to be considered below.
- Research question 4: Successful readers are clearly differentiated from less successful readers in terms of their *eye movements* when completing IELTS Academic reading test items onscreen.
- Research question 5: The data appear to show also that successful readers are indeed differentiated from less successful readers in terms of their *cognitive processing* while completing onscreen reading test (IELTS) items.

These findings match well with those in the earlier study (Bax 2013b).

8.2 Comparisons with previous research

Although the findings for Research questions 1 and 2 regarding global reading were inconclusive, for reasons reported above, the close alignment of this study with the previous project permitted additional insights into important differences between the two sets of findings in terms of Research questions 3, 4 and 5, as well as some interesting commonalities.

Whereas the earlier study found significant differences between the eye movement activity of successful and unsuccessful participants in test items 2, 3, 5, 7 and 10, no significant differences were found in the present study with regard to three of those items (2, 3 and 7). In the case of item 5, both studies showed the same result on one AOI, but this study also revealed a significant difference in another area as well; in the case of item 10, the present study identified an additional area of significant difference which was not present in the earlier study. In addition, this study identified significant differences in two items (4 and 11) which were not apparent in the earlier project.

One lesson to be drawn from this is that the patterns discerned in the earlier study were not confined to Malaysian readers, but are broadly consistent with reading patterns where other language and nationality readers are concerned. It also demonstrates the value of replicating such studies wherever possible with different nationality and language groups. Furthermore, if such studies were larger in scale, additional patterns might emerge from the eye-tracking data.

Notwithstanding, the results show again that eye-tracking is a powerful tool for understanding cognitive processing, when used with other methods, since this study was successful in identifying significant differences at a variety of levels of cognitive processing, as set out in Table 10, and also in terms of expeditious reading, to which we can now turn.

Levels of processing (see Table 1 for details)	Gloss	Relevant test item
Lexis: Word matching	Matching of identical word in question and in the text, as key to the answer	11 (*customer* in text)
Lexis: Synonym and word class matching	Matching of synonym or lexically related word in question and text	4 (*one type of/ one form of) in text and question* 10 (*housing* in text as near-synonym of *home* in the question)
Grammar/syntax	Focus on significant syntactic structure or other grammatical element as part of working out the answer	5 (*gives rise to* in the text)
Propositional meaning	Focus on elements, in question and/or text, required for constructing a propositional meaning essential to answering the question correctly	4 (*1986* in the text required for establishing correct propositional meaning) 5 (*1989* in the text, required for establishing correct propositional meaning)
Inference	Focus on elements, in question and/or text, required for constructing an inference essential to answering the question correctly	4 (*1986* in the text required for inferring and disambiguating the correct answer) 5 (*1989* in the text, required for inferring and disambiguating the correct answer)
Building a mental model	-	None, as this was not the focus of these IELTS test items
Understanding text function	-	None, as this was not the focus of these IELTS test items
Expeditious reading	Significant differences in terms of expeditious reading in two items, with unsuccessful students apparently unable to locate the site of a correct answer as effectively as successful students.	5, unsuccessful students spent significantly longer on larger chunks of text
		10, successful students were able to locate a smaller, particular part of a text and focus more expeditiously on it so as to extract the answer

Note: A 'gloss' is "a brief definition or synonym of unknown words provided in text in L1 or L2". (Nation, 2002, pp 174)

Table 10: Summary of findings for all items

8.3 Expeditious reading and metacognitive awareness

As shown in Table 10, in the case of two test items there were significant differences in terms of expeditious reading. Unsuccessful students were apparently poor at locating the site of a correct answer in the text, unlike successful students. This is clear from the fact that they spent significantly longer on larger portions of relevant text (item 5), while successful readers were able instead to locate specific area of the text, then focus on that section more expeditiously to determine the answer (item 10). This corroborated the findings with regard to expeditious reading in the previous study. Interview data showed that the reason for this was the use by successful readers of relatively conscious metacognitive strategies.

Unsuccessful students, however, appeared to use no such conscious strategies, apparently searching relatively randomly through the text to find the location of the answer.

This recalls the discussion above (section 4.1) deriving from earlier eye-tracking studies, in which it was suggested that more effective readers are likely to be more purposeful and focused in their reading. This could in future potentially be correlated with a possibly higher frequency of *return sweeps* and *corrective saccades* among successful readers, while less effective readers might be more speculative and unfocused, in turn making greater use of *backtracking*. Investigating this relationship between successful and unsuccessful second language readers under test conditions, with regard to their saccade activity, could be a potentially interesting area of future research, assuming a large enough cohort, and eye-tracking equipment capable of a high enough resolution.

However, even without such further saccade analysis, it is clear from the two studies that one central difference between successful and unsuccessful readers was the extent to which they operated strategically.

The differences in expeditious reading abilities, as revealed in the eye movement data, matched interview data which suggested that what differentiated successful readers from unsuccessful readers was the fact that most of them used pre-determined and conscious strategies systematically, whereas unsuccessful students for the most part appeared to be more aimless in their approaches. The strategies used by successful students varied widely from person to person, but in general what set them apart from unsuccessful students was the fact that they were distinctively purposeful and focused in their approach to the task at hand.

8.4 Other processes and strategies

An important aspect of this research, as also with the single nationality study, is that with a number of the test items (7 of 11), there were no significant differences between successful and unsuccessful participants in terms of eye-tracking evidence. As suggested in the previous study, it is probable that the main reason for this could be that successful readers make use of various combinations of knowledge, capacities and strategies which are not observable through eye-tracking. For example, it may be that the successful readers in this and the previous study had better lexical knowledge, better memory capacity or other qualities which were not reflected in eye movements.

This highlights the fact that, although eye-tracking has proven potential in analysing cognitive processing, it also has clear limitations. One implication of this for future research is that we should consider additional methods in addition to those used in this study, so as to try to elucidate a wider range of readers' cognitive operations, in order to understand how the various factors interact when participants read under test conditions.

Given the limited range of participants' IELTS scores, as noted in section 5.3, it would further be of value to research whether candidates from a wider ability range, in terms of candidates with higher and lower IELTS scores than those investigated here, for example, might make use of markedly different strategies in terms of eye activity, or whether the findings set out above would be replicated.

8.5 Implications for language test design and development

This study has strengthened the evidence base for using eye-tracking technology as a means of elucidating cognitive processing under test conditions, in conjunction with other methods such as those used here. Therefore, it is recommended that eye-tracking could usefully be employed in test validation of reading tests, in conjunction with other investigative approaches, to research cognitive validity and other areas.

The research also implies that item writers and testing bodies could in future usefully draw on the framework of cognitive processing levels developed by Khalifa and Weir (2009) so as to ensure that tests of academic English cover all the appropriate levels adequately.

There is still a need to research higher order cognitive processes in reading tests using eye-tracking, as this has not so far been attempted. In the case of IELTS, for example, it would be of value to research test items which target higher order processing such as 'building a mental model', and 'understanding text function' (see Table 1 above), if sufficient test items can be identified in a suitable test for analysis.

For reasons noted above, this research study was unable to come to conclusions concerning global reading as opposed to local reading, partly because of the difficulty of isolating suitable global reading items to investigate. It proved more straightforward to investigate local reading. An issue which this raises for reading tests in general is whether they perhaps focus too extensively on local reading, both in terms of quantity of items and also in test validation, to the relative neglect of global reading. One implication of this is that test developers could usefully reconsider the extent to which reading tests succeed in testing global reading, and how we might be able to establish the cognitive validity of such attempts.

A further implication for future language test design and development concerns expeditious reading. If, as the findings in this study show, the ability to read expeditiously is an important marker of successful as opposed to unsuccessful readers, then future reading test developers might well choose to give expeditious reading an even more central place in their specifications than they do currently. Given that expeditious reading can be assessed most effectively via computers, which can enforce timed reading components straightforwardly, it may well be that test designers will move even more comprehensively towards computerised modes of testing reading.

Finally, it is not inconceivable that in the course of time, it could be a plausible aim for the profession to construct a full, detailed and nuanced picture, through the careful use of eye-tracking technology and other methods, of readers' second-by-second mental activity as they read and complete test items. Such a model could make a significant contribution in future years to our understanding of test-taker behaviour and of the relative value of different reading test items.

REFERENCES

Bax, S, (2013a) 'Readers' cognitive processes during IELTS reading tests: evidence from eye-tracking', British Council, *ELT Research Papers,* pp 13–06, Available at: https://www.teachingenglish.org.uk/sites/teacheng/files/D069%20ELTRA%20report_FINAL_WEB%20ONLY.pdf

Bax, S, (2013b) 'The cognitive processing of candidates during reading tests: Evidence from eye-tracking' *Language Testing,* October 2013, vol. 30, no 4, pp 441–465

Bax, S and Weir, C, (2012) 'Investigating learners' cognitive processes during a computer-based CAE Reading test', *Cambridge Research Notes*, Cambridge ESOL Issue 47, February 2012, pp 3–14 http://www.cambridgeesol.org/rs_notes/rs_nts47.pdf

Baxter, G and Glaser, R, (1998) 'Investigating the cognitive complexity of science assessments', *Educational Measurement: Issues and Practices*, 17 (3), pp 37–45

Bertram, R, (2011) 'Eye movements and morphological processing in reading' *The Mental Lexicon* 6:1 (2011), pp 83–109

Buscher, G, Biedert, R, Heinesch, D and Dengel, A, (2010) 'Eye-tracking analysis of preferred reading regions on the screen' in *Proceedings of CHI Extended Abstracts 2010*, pp 3307–3312

Cambridge ESOL (no date), IELTS http://www.cambridgeesol.org/exams/ielts/index.html#acrd, Accessed 10 June 2012

Devi, S, (2010) *Investigating the academic reading of first-year undergraduates at a British university.* PhD thesis, University of Bedfordshire

Eger, N, Ball, L, Stevens, R and Dodd, J, (2007) 'Cueing Retrospective Verbal Reports in Usability Testing Through Eye-Movement Replay' in *Proceedings of HCI 2007, The 21st British HCI Group Annual Conference,* University of Lancaster, UK www.bcs.org/server.php?show=ConWebDoc.13300

Field, J, (2013) 'Cognitive validity' in A Geranpayeh and L Taylor (eds.) *Examining Listening,* pp 77–151. Cambridge University Press, Cambridge

Glaser, R, (1991) 'Expertise and assessment', in MC Wittrock and EL Baker (eds.) *Testing and cognition*, pp 17–30. Prentice Hall, Englewood Cliffs

Khalifa, H and Weir, C, (2009) *Examining Reading.* UCLES / Cambridge University Press, Cambridge

Nation, I. S. P. (2002).*Learning vocabulary in another language*. The Cambridge Applied Linguistics Series

Cambridge University Press.

Pallant, J, (2010) *SPSS Survival Manual,* 4th edition, McGraw Hill, Maidenhead

Rayner, K, (1998) 'Eye movements in reading and information processing: 20 years of research' in *Psychological Bulletin*, 124, 3, pp 372–422

Rayner, K, Pollatsek, A, Ashby J and Clifton, C, (2012) *The psychology of reading* (2nd ed.). Psychology Press, New York, NY

Reichle, E, Warren, T and McConnell, K, (2009) 'Using E-Z Reader to Model the Effects of Higher-Level Language Processing on Eye Movements during Reading', *Psychonomic Bulletin and Review*, 16(1), pp 1-20

Urquhart, A and Weir, C, (1998) *Reading in a second language: process, product and practice*. Longman, London

Wade, N, (2010) 'Pioneers of eye movement research' *Iperception*. 1(2): pp 33–68. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3563053/

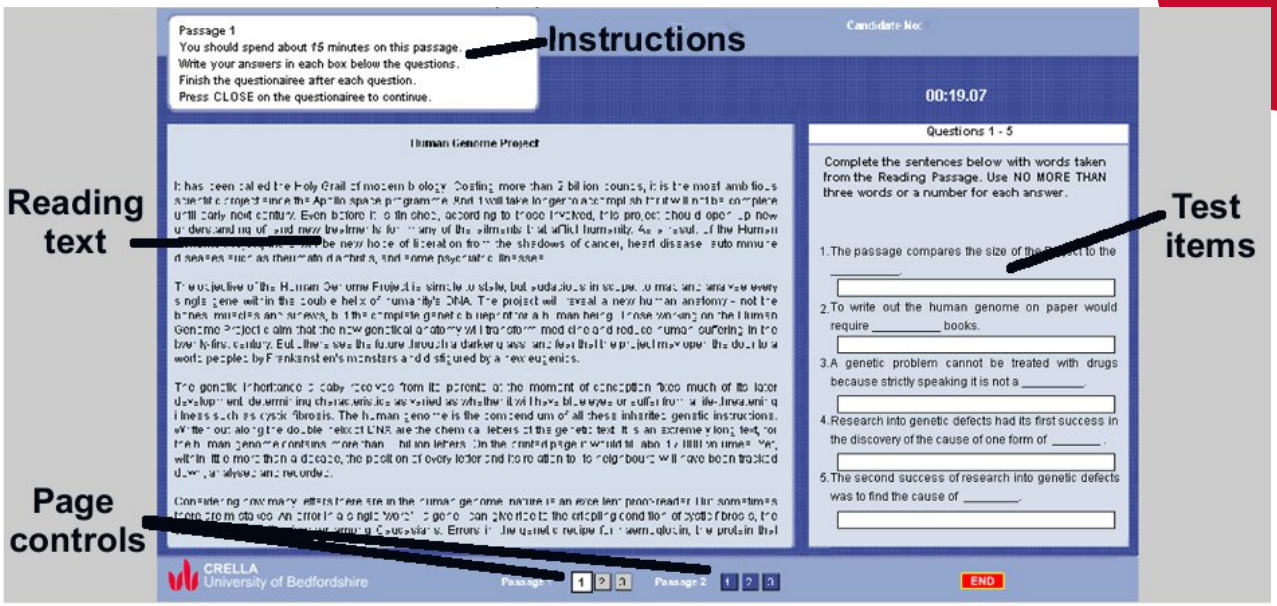
Weger, U and Inhoff, A, (2006) 'Attention and Eye Movements in Reading: Inhibition of Return Predicts the Size of Regressive Saccades' in *Psychological Science*. March 2006; 17(3), pp 187–191

Weir, C, Hawkey, R, Green, A, Unaldi, A and Devi, S, (2012) 'The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university' in L Taylor and C Weir (eds.), *IELTS Collected Papers 2: Research in Reading and Listening Assessment*, Volume 2, pp37–120, Cambridge University Press

Weir, C, Hawkey, R, Green, T and Devi, S, (2009) 'The cognitive processes underlying the academic reading construct as measured by IELTS', *IELTS Research Reports* Volume 9, 4: pp 157–189, British Council, London and IELTS Australia, Canberra

Weir, C and Khalifa, H, (2008) 'A cognitive processing approach towards defining reading comprehension' *Cambridge ESOL Research Notes,* Issue 31, February 2008, pp 2–10

APPENDIX 1: LAYOUT OF THE TEXT AND TEST ITEMS



APPENDIX 2: *TEST ITEMS*

a) Part 1: Gap-fill questions

1. The passage compares the size of the Project to the _____
2. To write out the human genome on paper would require ____books.
3. A genetic problem cannot be treated with drugs because strictly speaking it is not a_____
4. Research into genetic defects had its first success in the discovery of the cause of one form of _____.
5. The second success of research into genetic defects was to find the cause of _____.

b) Part 2: Matching task

List of Biometric Systems	List of users to be matched by the test-taker while reading
A. fingerprint scanner B hand scanner C body odour D voiceprint E face scanner F typing pattern	6. sports students 7. Olympic athletes 8. airline passengers 9. welfare claimants 10. home owners 11. bank customers

APPENDIX 3: ANALYSIS OF EACH TEST ITEM IN TERMS OF ANTICIPATED COGNITIVE PROCESSING

Item	Level of text	Type of processing required	Target in question item - AOI	Target in the reading texts - AOI
2.	Within one sentence	Lexical, Synonymy	*Books* *On paper*	*Printed page* *7000 volumes*
3.	Within one sentence	Lexical matching Syntactic parsing	*Drugs* *Genetic defects*	*Drug* (only occurrence in text) *Single-gene disorders* *disease*
4.	Across sentences	Inference (*First)* Lexical, Synonymy	*First success* *One form of*	*In 1986 (compared to 1989 later)* *One type of* *Muscular dystrophy*
5.	Across sentences	Inference (*dates*) Lexical, Synonymy	*Second success* *The cause of*	*In 1989 (compared to 1986 earlier)* *Gives rise to* *Cystic fibrosis*
6	Within one sentence	Lexical matching Synonymy	*Sports students*	*Students, athletic* (para A)
7.	Within one sentence	Lexical matching	*Olympic athletes*	*Olympic, athletes* (para E)
8.	Within one sentence	Lexical matching Synonymy	*Airline passengers*	*Passengers, airport* (para F)
9.	Within one sentence	Lexical matching	*Welfare claimants*	*Welfare, welfare payments* (para D)
10.	Within one sentence	Synonymy	*Home owners*	*Housing (para A)*
11.	Within one sentence	Lexical matching	*Bank customers*	*Customers, Bank (para A)*