White space in assessment materials – ‘space to think’ or a ‘waste of space’?

Charlotte Stephenson[a] and Dr Bryan Maddox[b]

[a] *AQA, Manchester, UK*

[b] *Assessment MicroAnalytics™, University of East Anglia, Norwich, UK*

Abstract

In this paper, we explore the significance of white space in test paper design in the context of high-stakes general qualifications in England. Exams should assess a student’s ability in relation to a given construct (e.g. chemistry). However, the influence of text layout on performance may threaten the construct validity of assessments. This research explored the effects of white space in question papers on cognitive processing – using eye-tracking methods – and respondent perceptions. The eye movements of 32 students (aged 15–16 years) were tracked as they completed two abridged AQA GCSE Chemistry papers: one with restricted spacing and one with enhanced white space. Eye-tracking data showed that respondents took longer to complete questions with enhanced white space but also made more careful observations of the question content. Conversely, restricted white space was associated with shorter response times and more frequent rereading of item content. Interview data revealed that students preferred papers with enhanced white space, reporting that the additional space made them feel calmer and that the papers were easier to read. These findings therefore suggest that the amount of white space in assessments not only impacts respondent preferences but also leads to measurable differences in assessment response processes. We discuss the implications of these findings for increasing validity in assessment design.

Introduction

In this paper, we describe a mixed-methods research project that explored the significance of white space in test paper design, specifically in high-stakes general qualifications (GCSEs) in England. Little is known about the effects of test paper layout, particularly the configuration of white space, on academic performance. To our knowledge, the effects of white space in question papers that contribute to high-stakes general qualifications in England have not yet been examined. We discuss the degree to which white space in assessment design may represent a source of construct-irrelevant variance in test scores, by drawing on our work that looked at the eye movements of students as they complete questions papers with varying proportions of white space. We consider the implications that white space has on the cognitive processes of students, and also explore students' perceptions of white space.

The paper has four main sections. In this first section, we explain why this is an important issue for investigation and introduce the emerging literature on test paper layout. In the second section, we set out the methods used to investigate our research questions. In the third section, we explain the main eye-tracking findings and outline the qualitative findings, which we elucidate using students' words. We discuss these findings in relation to existing literature and the research questions posed. We also evaluate their dependability and explain their implications for assessment design. In the fourth and final section, we offer some conclusions that can be drawn from the research and some recommendations for assessment designers.

Construct validity

In the context of high-stakes general qualifications, examinations should assess a student's ability in relation to a given construct (e.g. chemistry) while eliminating as much noise as possible. One potential source of construct-irrelevant variance that might interfere with the accuracy of educational measurement could be the ways in which information is presented and formatted in exam papers. More specifically, teachers have suggested that the amount of white space presented in GCSE science question papers could affect their accessibility. If this is the case, then instead of assessing a student's knowledge and understanding of chemistry, the examination may be assessing something else – i.e. a student's ability to read information presented in a certain way. This could call into question the validity of the qualification.

Moreover, exam layout has been identified as a stressor for students that may lead them to demonstrate uncharacteristic behaviour in examination situations (Weir, 1993). Indeed, many researchers have emphasised that assessment materials must be legible and accessible to ensure that they appropriately measure the skills intended (Davis, 1993; Hughes, 1989; Weir, 1993; Zimmerman et al., 1990).

White space

White space refers to the amount of inked surface relative to blank paper in printed materials; this includes space between letters, words, lines of text and headings. Print design experts recommend the generous use of white space to aid the legibility of text (Tinker, 1963; Wilson, Pfister & Fleury, 1981), with some claiming that it can provide a logical structure for the reader to follow (Hoener, Salend & Kay, 1997) and reduce visual stress (Hughes & Wilkins, 2002).

Research investigating the effects of text layout in assessment materials is scant; however, some understanding can be drawn from the literature on reading comprehension. Lonsdale (2007) found that the layout of assessment materials can affect test performance: reading comprehension scores were found to be higher when text passages and question and answer

sheets had more white space. Participants in the study reported preferring texts and question and answer sheets with more 'interlinear space, space between paragraphs, space within the list of questions, space within the list of answers' (Lonsdale, 2007, p. 30). However, to our knowledge, the effects of white space on performance in high-stakes tests have not yet been explored. The significance of white space in assessment materials must be understood if we are to maximise validity in assessment design in this context.

Eye tracking in assessment

Eye tracking provides detailed information on respondent gaze in assessment that can be used to inform test design and the analysis of test performance and validity (Bax, 2013; Oranje et al., 2018; Maddox et al., 2018). Eye-tracking studies in assessment are typically small-scale, investigating user experience in computer-based assessments. Their focus is usually on aspects of scene perception: how respondents engage with areas of interest (AOI) within test items. This contrasts with more detailed, high-frequency studies of eye movement in reading research (e.g. Rayner, 2009).

Eye tracking can play an important role in test development and validation (Ercikan & Pellegrino, 2018; Zumbo & Hubley, 2018), especially when combined with other sources of information such as computer log files and cognitive interviews. The indicators of variation in assessment performance typically considered in assessment research include scan-path patterns, the frequency and duration of fixations, and the frequency of rereading (re-visits) within an item and between items (e.g. looking back at a previous item).

Rationale

There is a paucity of empirical research regarding the effects of white space in assessment materials. This research was intended to add to the literature by establishing the effects of variations in white space in GCSE Chemistry question papers on aesthetic preferences and cognitive processing.

In this study, we combined eye-tracking observations with post-assessment interviews to investigate how different amounts of white space in exam papers influence respondent preferences and cognitive processes.

In order to improve ecological validity, our study had two novel aspects:

1. We took eye tracking out of the lab, to observe assessment response processes in students' school environment. This was informed by a trend in observing assessment practices 'in vivo', as they take place in real-world contexts (Zumbo, 2017; Maddox & Zumbo, 2018).
2. We used eye tracking to capture response processes as students completed paper-based GCSE exam scripts, applying eye-tracking techniques that are normally associated with computer-based modes of assessment.

We had two research questions:

RQ1: How does the amount of white space in GCSE Chemistry assessments affect students' perceptions and preferences of question paper design?

RQ2: How does the amount of white space in GCSE Chemistry assessments affect students' cognitive processes?

In addition to these questions, we explored the hypothesis that while additional white space may be aesthetically pleasing to respondents (RQ1), it would not produce differences in cognitive processes or associated response times.

Method

Design

A within-subjects design was used in which two abridged versions of GCSE Combined Science (Trilogy) Chemistry papers were mocked up using the live papers from June 2018. There were two versions of each paper (i.e. four papers in total): one with restricted spacing and one with enhanced white space throughout. Each student saw both papers and both layouts in various combinations. A Graeco-Latin square design was used to balance the combination of each test paper with each layout, which controlled for order effects and effects of test demand. Each of the tests (Paper 1 and Paper 2) was seen an equal number of times in each layout (enhanced and restricted), and each paper was seen first and last an equal number of times across the sample. To achieve this, the sample was allocated into four subgroups, with eight students in each:

1. Paper 1 with restricted spacing, and then Paper 2 with enhanced white space
2. Paper 1 with enhanced white space, and then Paper 2 with restricted spacing
3. Paper 2 with restricted spacing, and then Paper 1 with enhanced white space
4. Paper 2 with enhanced white space, and then Paper 1 with restricted spacing.

Students were given 12 minutes for each question paper to ensure they could complete both papers and an interview within standard allocated lesson time. Each paper was worth 11 marks and the time allocated per mark was proportional to that of a full live GCSE Combined Science (Trilogy) Chemistry question paper.

Participants

The participants were 32 GCSE science students at the start of year 11 (aged 15–16 years), none of whom had a reading disability. Diversity of ability was sought to allow for an exploration of differences in the effects of layout according to student ability.

Recruitment

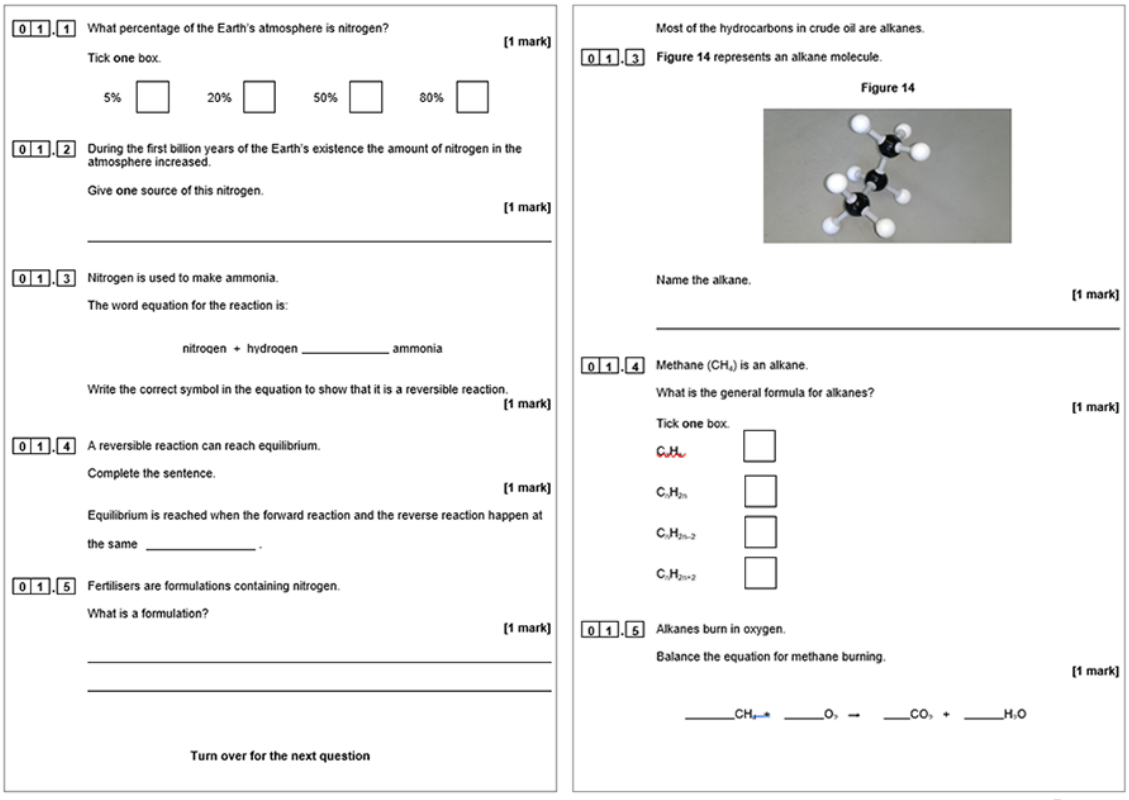
Large comprehensive schools in the north of England, with which AQA's Curriculum team have a relationship, were approached. One willing school was selected and 60 eligible students were given details of the study; 32 of these students opted to participate.

Materials

Two abridged versions of GCSE Combined Science (Trilogy) Chemistry papers were mocked up using items from the Summer 2018 live papers, referred to here as Paper 1 and Paper 2. Two versions of each paper were produced: one with enhanced spacing (Version A) and one with restricted white space throughout (Version B). Items across Paper 1 and Paper 2 were matched as closely as possible to ensure that the papers had a similar level of difficulty. This was achieved using the item facility indices, based on students' performance in summer 2018 (a facility index gives a measure of an item's difficulty, calculated by dividing the question mean mark by the maximum number of marks available). Only learning content that had been covered by the school was included in the materials.

For all papers, the layout was based on the standard AQA formatting for GCSE science papers, using 11-point type size in Arial typeface. Single line spacing was used in the restricted white space papers with a maximum of two lines given between questions (see Figure 1 for an example of a page from Papers 1B and 2B). Line spacing was set at 1.5 lines in the enhanced papers to increase the amount of white space throughout (see Figure 2 for an example of a page from Papers 1A and 2A). The amount of space between questions varied in the enhanced papers and was maximised depending on the amount of space available but typically around five lines. For all papers, as in reality, the arrangement of items and, consequently, the spacing between items was sometimes driven by pragmatic considerations depending on item length in order to fit items belonging to the same question or topic on the same page and to ensure new questions and topics started on a new page.

All four papers were printed in 8-page A4 answer booklets, printed on two sheets of A3 paper, double-sided and folded in half and stapled down the centre. The variations in white space meant the number of blank pages at the back of the booklet varied, while the size of the booklet was consistent across all four versions.



0 1 . 1 What percentage of the Earth's atmosphere is nitrogen?

[1 mark]

Tick one box.

5% ☐ 20% ☐ 50% ☐ 80% ☐

0 1 . 2 During the first billion years of the Earth's existence the amount of nitrogen in the atmosphere increased.

Give one source of this nitrogen.

[1 mark]

0 1 . 3 Nitrogen is used to make ammonia.

The word equation for the reaction is:

nitrogen + hydrogen _____ ammonia

Write the correct symbol in the equation to show that it is a reversible reaction.

[1 mark]

0 1 . 4 A reversible reaction can reach equilibrium.

Complete the sentence.

[1 mark]

Equilibrium is reached when the forward reaction and the reverse reaction happen at the same _____ .

0 1 . 5 Fertilisers are formulations containing nitrogen.

What is a formulation?

[1 mark]

Turn over for the next question

Most of the hydrocarbons in crude oil are alkanes.

0 1 . 3 **Figure 14** represents an alkane molecule.

Figure 14

Name the alkane.

[1 mark]

0 1 . 4 Methane (CH_4) is an alkane.

What is the general formula for alkanes?

[1 mark]

Tick one box.

C_nH_n ☐

C_nH_{2n} ☐

C_nH_{2n-2} ☐

C_nH_{2n+2} ☐

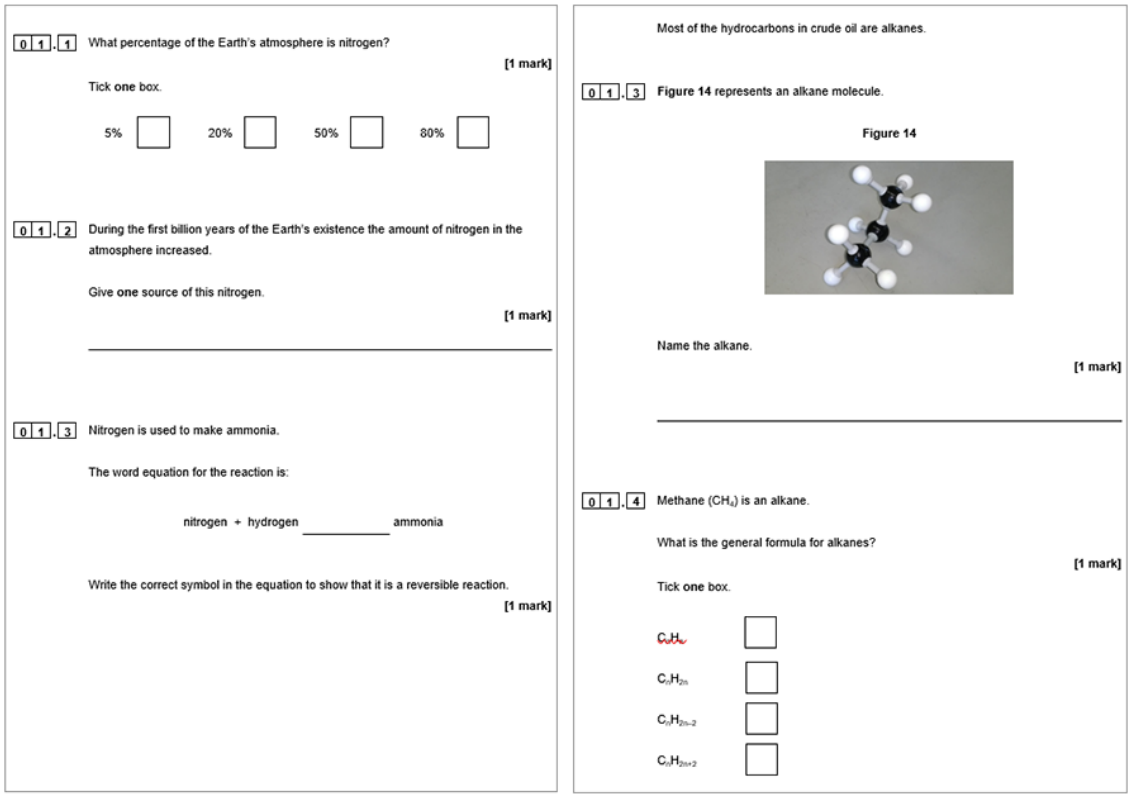
0 1 . 5 Alkanes burn in oxygen.

Balance the equation for methane burning.

[1 mark]

_____CH_4 + _____O_2 → _____CO_2 + _____H_2O

Figure 1 Extracts from Paper 1B (left) and Paper 2B (right)



0 1 . 1 What percentage of the Earth's atmosphere is nitrogen?

[1 mark]

Tick one box.

5% ☐ 20% ☐ 50% ☐ 80% ☐

0 1 . 2 During the first billion years of the Earth's existence the amount of nitrogen in the atmosphere increased.

Give one source of this nitrogen.

[1 mark]

0 1 . 3 Nitrogen is used to make ammonia.

The word equation for the reaction is:

nitrogen + hydrogen ⟶ ammonia

Write the correct symbol in the equation to show that it is a reversible reaction.

[1 mark]

Most of the hydrocarbons in crude oil are alkanes.

0 1 . 3 **Figure 14** represents an alkane molecule.

Figure 14

Name the alkane.

[1 mark]

0 1 . 4 Methane (CH_4) is an alkane.

What is the general formula for alkanes?

[1 mark]

Tick one box.

C_nH_n ☐

C_nH_{2n} ☐

C_nH_{2n-2} ☐

C_nH_{2n+2} ☐

Figure 2 Extracts from Paper 1A (left) and Paper 2A (right)

Procedure

First, the eye-tracking glasses were calibrated to ensure eye movements were recorded accurately. Students then completed both question papers (12 minutes provided for each) wearing the eye-tracking glasses. While one researcher was working with the students on the eye-tracking experiment, another researcher conducted post-task interviews (approximately 10 minutes each) with students in a quiet private room to obtain their perceptions of the paper layout and item design. The interview addressed students' attitudes towards the papers, ascertaining their first impressions of the papers, how difficult they found the papers and their perceptions of the layout of the papers. During the interview, the students were encouraged to consult the question papers they had taken.

Piloting

The experimental procedure was piloted with a selection of students in the school to ensure the methods were appropriate and the time allocated sufficient. No changes were made to the procedure at this stage and the data was deemed appropriate for inclusion in the analysis.

Measures

Eye-tracking glasses were used to collect various data on cognitive processing, including:

1. item response processes, with data on respondent gaze scan-paths and the frequency and duration of fixations
2. the frequency of rereading, as re-visits to items and other AOIs
3. response times, for the time taken to read and answer questions.

The question papers were marked by an AQA examiner. The test scores were used to compare performance between the tests. However, because of the small sample size we did not draw conclusions about the impact of white space on test scores.

Equipment, software and data quality

We used a pair of SMI ETG 60hz wearable eye-tracking glasses with a sampling rate of 60hz. For the data analysis we used SMI Be-Gaze (3.7) Software. This enabled us to aggregate and compare gaze data on selected reference images and AOIs, and to select and observe behaviour for subgroups.

The software enabled us to observe scan-path videos, and to produce saccade-oriented data visualisations, including heat maps, sequence charts and AOI statistics. We were also able to use scan-path videos to capture accurate data on item response times (data that one would normally associate with log files in computer-based tests). For our data analysis (subject to page turning), we considered individual exam questions as our reference image and primary unit of analysis.

During the data collection, the test administrator sat adjacent to the student, observing the eye-tracker video in real time on a laptop screen. This enabled us to observe the eye-tracker performance and the quality of data, and to communicate with the students as necessary. We provided 10 minutes prior to data collection to complete the manual (3-point) calibration process, to allow each student to become familiar with wearing the eye-tracker glasses, and to fine tune the eye-tracker performance. The quality and robustness of eye tracking can be reduced due to factors such as eye and nose shape, the use of contact lenses, and mascara (Holmqvist et al., 2011). In our study, within the school context, around 20 per cent of the sample was affected by some of these problems.

There is a trade-off between the benefits of using eye-tracking glasses and screen-mounted eye trackers in assessment research. Screen-mounted eye trackers only capture screen-based gaze data. In contrast, eye-tracking glasses capture more holistic gaze data on response behaviour such as social interaction, gaze aversion, calculator use and the use of pen and paper. However, eye-tracking glasses are more invasive and less accurate than screen-mounted eye trackers. They do not enable high-frequency, saccade-oriented eye tracking that would provide deeper insights into rapid scanning behaviours and peripheral vision (see Rayner, 2009; Niefind & Dimingen, 2016; Leube, Rifai & Wahl, 2017; Struckelj & Niehorster, 2018).

Interview analysis

Eye-tracking analysis

The eye-tracker data was analysed using SMI Be-Gaze (3.7) software. AOIs were created around the key features of items, such as instructions, stimulus diagrams and text. The residual area outside the AOIs was classified as white space. The software enabled us to capture and visualise respondent scan-paths, and the order, frequency and duration of fixations (including AOI returns and skips). This follows established methodology in eye-tracking analysis (Holmqvist et al. 2011).

Qualitative analysis

Thematic analysis was used to analyse the transcribed interview data and identify dominant themes in students' perceptions of assessment design. Braun and Clarke's (2006) guidelines directed the analysis and Quirkos was used to classify, sort and arrange information (Quirkos, 2017). The analysis was inductive: driven by the data and not guided by existing theory (Hayes, 2000).

Results and discussion

In this section, we report the findings of the eye-tracking experiment and the qualitative findings from the interviews. We discuss the relevance of the findings in relation to existing literature and consider potential avenues for further research. Finally, we evaluate the credibility of this work.

Interview data (RQ1)

Before discussing the themes that dominated the interviews with students, it is important that we first outline a few observations relating to the students' perceptions, as they should be borne in mind when considering the themes. These observations pertain to students' first impressions of the papers and their preferences.

First impressions

The majority of learners did not notice the differences in layout of the papers at first. However, once we invited the students to visually compare the papers side by side, they began to draw out differences. This is typified by Participant 1:

> I just realised that, I didn't realise while I was doing it but there's quite a lot of gaps at the bottom, and here there's not a lot of gaps at all. (Participant 1)

This suggests that the effects of text layout may be unconscious for the learners to some degree, as they focus on answering the questions. Participant 3 said:

> I don't really think about how it's all laid out when I'm doing it. Just read the questions. (Participant 3)

Learner preferences

Many learners found the amount of space to be acceptable in both papers:

> Interviewer: On Paper 2B, you said you think there's less space. Do you think there's enough space?
>
> Participant 9: Yeah.

However, distinct preferences emerged; students typically preferred the papers with enhanced white space. This was prominent for students who saw the 1B/2A combination compared with the 1A/2B combination.

> Paper 2[A] was fine. I really liked the space, how spaced out it was. But Paper 1[B], I really didn't like it. (Participant 16)

It seems the impact of the variation in white space between these two papers was more extreme.

Indeed, these findings further support prior research, which has demonstrated that learners tend to prefer assessment materials with generous amounts of white space (Chelesnik, 2009; Lonsdale, 2007).

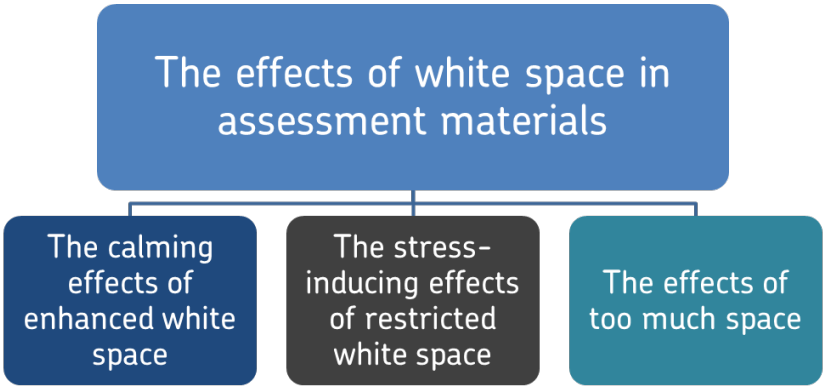


Figure 3 Themes derived from the student interview data

Students' experiences were dominated by three main ideas regarding how the amount of white space affected their perceptions (see Figure 2). The first related to the calming effects that white space had on the students' assessment experience, the second theme concerned the stress-inducing effects of restricted space and the third emphasised the effects of too much space. The meanings behind the three themes are elucidated below in turn, and exemplified using the students' words. The themes are interpreted in relation to existing literature and the research questions posed, their credibility is evaluated and their implications for assessment design elucidated.

The calming effects of enhanced white space

The students talked about several ways in which the enhanced white space had a calming effect on their assessment experience.

More space in the assessment makes learners feel calmer

The extra space in the enhanced papers was referred to as being calming for students; words like 'relaxed', 'comfortable' and 'calm' were used by learners to describe their experience of completing the papers.

> With this one, because it's so spaced out, it's almost calming in a way. (Participant 16)

> In Paper 1B, the first page, when you look at it straightaway you think wow. But once you're at Number 2A, it was a little more relaxed, as in there's more space. (Participant 30)

Some referred to the space affecting their confidence and allowing them to believe they could complete the paper.

> If there's more space it gives you, I don't know, it just makes you feel more confident about the rest of the paper. Because you're like OK, if I can ease myself into this, and the structure of paper isn't that much, like there's not many questions, I think I can do that. (Participant 2)

Some learners reported feeling more relaxed about the time they had to work on the questions when the space was more generous.

> When there's a question on the page, one question, it sort of tells you what, it's one question per page … I don't know, it just gives me a sense that I've got more time to answer a question, and maybe work on it more. (Participant 22)

More space makes the assessment seem easier

The calming effect of the enhanced white space seemed to be associated with students' perceptions of demand, making papers feel 'easier'.

> But Paper 2[A], again it was calming, and I saw it and I was like I know I can answer all of these. But Paper 1[B] felt intimidating. (Participant 16)

> I think more space makes it seem easier to answer. (Participant 7)

The students talked about how the enhanced white space made the test easier to read, follow and process.

> I think because there's space between the questions it was a lot easier to process. (Participant 19)

Some learners talked about how space helped them to answer questions successfully by allowing space for working out.

> Well they give you spaces to jot things round the paper itself. So that helps. (Participant 12)

The paper was described by some students as looking 'friendly' and 'nicer' than the restricted paper. Based on these descriptions, students' experience of completing the enhanced papers was seemingly more enjoyable.

The finding that students felt calmer when completing the enhanced white space version seems to be upheld in the eye-tracking data; response times were slower for questions with enhanced space and accuracy increased. If students feel calmer and more able to take their time answering the questions, their reading accuracy could be augmented. While our sample was not large enough for differences in test performance to be detected between the two paper layouts, it seems reasonable to suggest that enhanced reading accuracy might impact upon test performance on a larger scale. Indeed, Lonsdale, Dyson and Reynolds (2006) found that test layout affected reading comprehension scores and concluded that, based on participants' comments, the interlinear space and separation of paragraphs were the typographic features most likely to cause the effect of text layout on performance.

In addition, students reported finding the white space on the page helpful. Research suggests that test performance is impeded when no space is provided for working out (Scott & Webb, 1979).

The stress-inducing effects of restricted white space

In contrast, students reported that restricted white space in assessment materials had a stress-inducing effect.

Less space makes the assessment seem harder

Students reported feeling that the papers with less white space were more difficult, some referring to this layout as ‘confusing’ and ‘complicated’.

> [The questions are] all really close together. I think it’s a bit harder to read everything and understand it if it’s like that. (Participant 4)

Less space makes the assessment feel stressful

The perception of increased demand appeared to have a stress-inducing effect on students.

> When you’re first having a look at it you think, your mind thinks there’s no chance I’m going to get it, there’s too much words. But when it’s like less and it’s more spread out you feel comfortable, your mind’s thinking I can do it, I can do it. When there’s too much you think, your head just thinks I can’t do it. (Participant 30)

> All the questions are really close together, which can again seem a bit overwhelming. (Participant 2)

Some learners reported that the stress associated with the restricted space made them feel rushed.

> When there’s so much on a page you kind get a bit worked up, like you need to answer them quickly or you need to get through them. (Participant 22)

Less space in the assessment feels constricted

When talking about the papers with less space, learners frequently used words associated with feelings of constriction. Terms like ‘squashed’, ‘cramped’, ‘crammed’, ‘condensed’, ‘crowded’ and ‘packed’ were used to describe the spacing of words and/or questions in the restricted papers. This seemed to make reading more difficult.

> Whereas 2B it was quite cramped, quite all crushed together which is probably why I did miss out some of the hints given. (Participant 24)

The finding that students felt more stressed and rushed when white space was restricted corresponds with the eye-tracking data. It is possible that the perceived difficulty and consequent stress caused students to feel rushed, resulting in faster item response times. Typographic design determines how well readers assimilate the information contained within the text (Rotter, 2006). Indeed, Hughes and Wilkins (2000) assert that ‘text can be visually stressful and thereby affect reading ability’ (p. 322). The observation that reading was less accurate and re-visits were more frequent in the restricted version may explain why students found the restricted papers to be more difficult. However, cause and effect here is difficult to determine.

The effects of too much space

Some students, however, did communicate that there was a limit to the benefits of enhanced white space in assessment materials, reporting counterproductive effects of too much space.

Too much space makes the question/paper look bigger

A small number of students mentioned that too much space in an assessment can make the question or the whole paper appear bigger and more daunting.

> I liked the fact that it had one question on one paper. But when you look at it you're going to think it's too much, and too much questions… when I got handed this it felt thick. So straightaway you think oh it's a lot. (Participant 30)

> So this question. It's all really spaced out, but there's only actually one thing that you need to write as an answer. So it makes the question look a lot bigger than it actually is. (Participant 3)

Too much space wastes paper

Some students expressed concerns about waste if too much space is incorporated into an assessment.

> I think this was a waste of space, there's nothing going on. (Participant 26)

> I think there could have been a bit more space between questions, but like you don't want to use an excessive amount of paper. (Participant 19)

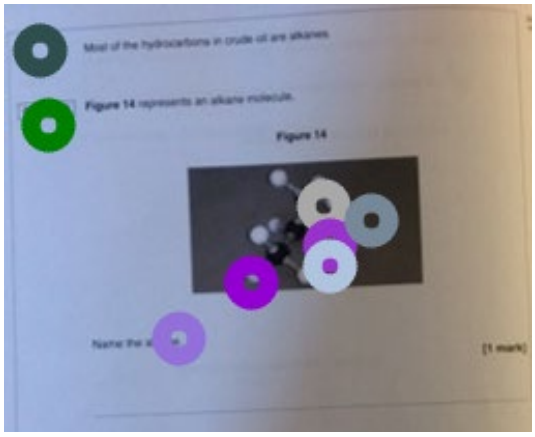
While learners reported a preference for more space in assessments, there were concerns expressed that this could become unreasonable or counterproductive. There are cost and environmental considerations that place limits on the amount of space and paper that can reasonably be used in printed materials (Bradshaw & Johari, 2002).

Eye-tracking data (RQ2)

From the eye-tracking data, we observed that the respondents engaged well with the assessment tasks. They answered the questions in the order in which they appeared in the exam paper, although they did not always read item content in the order in which it was presented. They sometimes jumped ahead to the questions, or to information contained in diagrams, before returning to read the item instructions.

Each student had 12 minutes to complete each paper. They normally worked fast with a sense of urgency, completing the exam paper in around half the time provided, before returning to review their work and to complete any unanswered questions. It is not clear that this strategy was beneficial, but it seems to reflect their approach to completing exam papers.

In most cases the gaze 'landing spots' (the locations of the initial fixations) were the question numbers at the top left of each page. However, in some cases, attention was initially captured by diagrams presented in the item (see Slide 1 below).

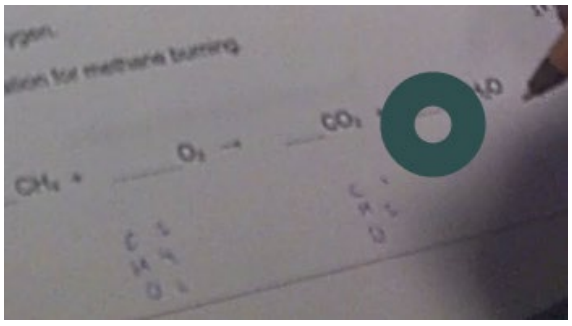


Slide 1 **Initial fixations of eight participants on the question numbers and the diagram**

It is clear from the observations of eye-tracker scan-path data that, for both versions of the papers, respondents read the content of more difficult items more frequently, often rereading some sections of the text several times before providing their answer.

Dimensions of white space

Before we can consider the impacts of white space on response processes, we should recognise the different dimensions of white space. White space can occur within an item, between items, and outside items (e.g. in margins). For example, Paper 1 (Questions 1.1–1.5) had visibly less space and offered little white space within and between the items in the restricted white space version. While in Paper 2 (Question 1.5), the restricted white space version left little space outside the item at the foot of the page for the respondent to work out their answer. The exam paper instructs the students 'Do not write outside the box'. In this case, then, restricted white space limits students' space to write and could adversely impact on respondent performance (see Slide 2).

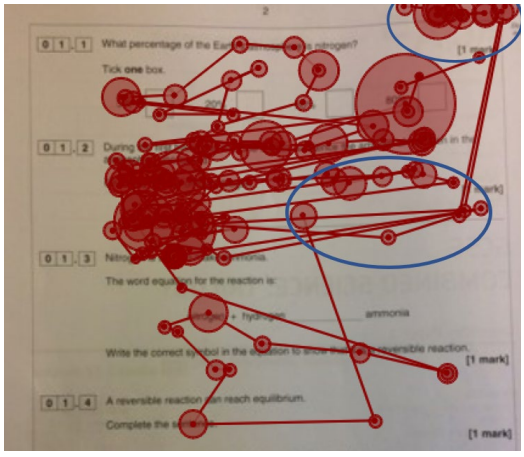


Slide 2 **Restricted space to write at the foot of the page (Paper 2: Question 1.5)**

Fixations in areas of white space

In addition to providing additional space to write, the data suggests that the students may have used white space as 'space to think'. Respondents made repeated fixations in areas of white space within and outside the test items. Fixations in areas of white space were considerably more frequent in the papers with enhanced spacing.

The frequency of these behaviours suggests a feature of cognition rather than errors in eye-tracker alignment (see Slide 3).



Slide 3 Fixations in areas of white space within and outside the test item (circled)

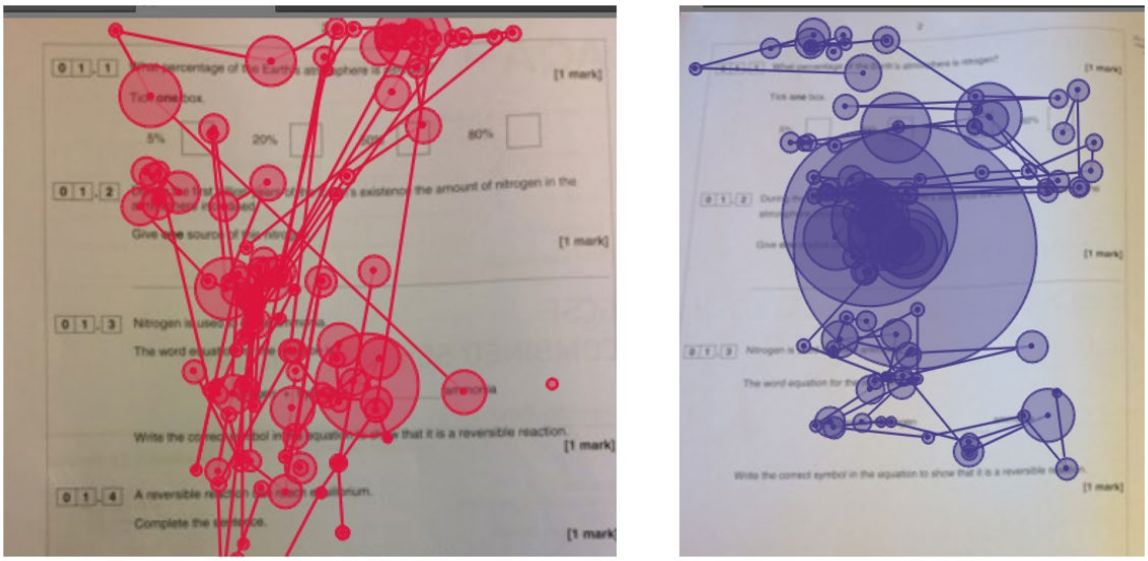
It has been argued that white space makes text easier to read as it provides areas for readers to rest their eyes and gives the brain a chance to process information (Baker, 2001; Bradshaw & Johari, 2002; Cubberley, 1991). This view is supported by the literature on gaze aversion, which suggests that people sometimes divert their gaze away from a stimulus as a strategy to reduce cognitive load (see Glenberg et al., 1998; Doherty-Sneddon & Phelps, 2005; Seli et al., 2016). Contrasting hypotheses either interpret gaze aversion as 'deliberate, intentional thought' (Seli et al., 2016), or as disengagement or 'mind wandering' (Varao-Souza et al., 2017; Bixler & D'Mello, 2016; Seli et al., 2016). In our study, either explanation is plausible. However, the timing and location of fixations in areas of white space suggests it was a cognitive strategy (deliberate or otherwise), rather than mind-wandering. Its key features are its short duration (fractions of a second) and the close proximity of gaze aversion to test item content (i.e. within the item or on the margins of the paper). Due to sample size, we are not able to test these hypotheses in terms of any significant performance-related differences in test scores.

The impact of white space on response processes

In our research, response times were on average slightly longer for the papers with enhanced spacing than for the papers with restricted white space (13 seconds longer for Paper 1 and 15 seconds longer for Paper 2). We can attribute that to longer fixations in areas of white space with the enhanced versions.

We did observe variation in response times within each of the two different formats. As a result, it is possible that observed differences in response times between the two formats could be the result of confounding factors, or some idiosyncratic features of the sample. However, the eye-tracking data reveal substantive differences in gaze patterns that help to explain the differences in response times.

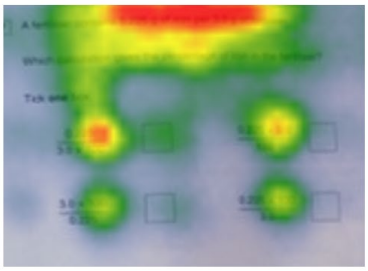
Overall, the number of re-visits between the items in the restricted white space paper was around twice that of the enhanced white space paper. Whereas the number of fixations in areas of white space was over four times greater in the enhanced version than in the restricted version (see Table A1 in the appendix). Slide 4 (below) illustrates the different gaze paths and fixations of respondents for the two versions of Paper 1, in relation to Questions 1.1–1.5. For those questions, the response times were on average six seconds longer for the version with enhanced spacing. This illustrates how the amount of white space impacts on response processes.



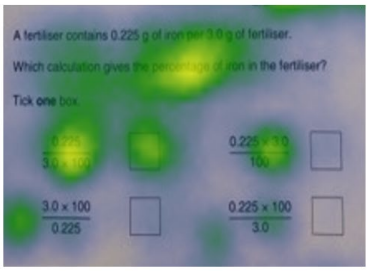
Slide 4 **Illustrative examples of scan-paths for restricted (left) and enhanced (right) white space**

White space and attention to detail

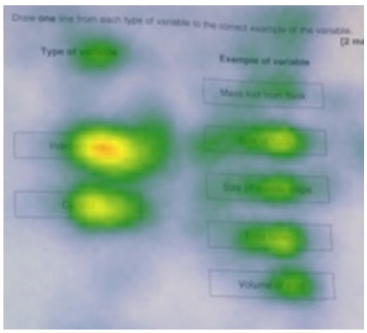
A final observation from the eye-tracking data is that in some of the test items, the respondents seem to have paid closer attention to the detail of the item in the enhanced papers, with more lengthy reading of key item content. Slide 5 provides some ‘heat map’ examples of how the attention to detail varied across the enhanced and restricted white space formats.



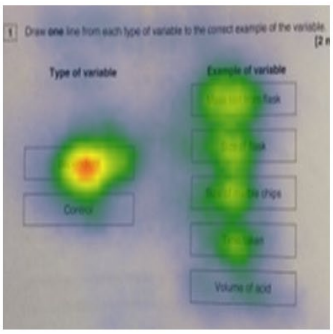
Enhanced



Restricted



Enhanced



Restricted

Slide 5 Heat map comparisons of attention to the detail of test items

This observation is supported by the work of Hoener, Salend and Kay (1997) who assert that increased line spacing allows children's eyes to scan the correct line of text more easily, which should decrease the number of times children reread or skip lines of text. In our work, the observations are not sufficient to establish a robust relationship between white space and attention to detail. However, they suggest that it would be worth investigating this in future research.

The eye-tracking evidence suggests that we should reject our negative hypothesis that while additional white space may be aesthetically pleasing to respondents, it would not produce differences in cognitive processes or associated response times. The preferences of the students in the post-assessment interviews were supported by empirically observed differences in response processes across the enhanced and restricted white space designs, with differences observed in the item response times, the frequency of rereading (re-visits), and the evidence around attention to detail.

Conclusions and recommendations

The purpose of this research was to investigate the impact of enhanced white space in GCSE science papers. The research was motivated by pragmatic concerns about test paper design. We conducted the research 'in vivo' in a diverse school in England to investigate the views and responses of GCSE students.

The participants in our study reported a preference for the exam papers with enhanced space, stating aesthetic and substantive reasons for their preference.

The students reported that the enhanced white space appeared easier and made them feel calmer and less rushed. Their stated preferences in post-assessment interviews were also supported by empirical eye-tracking data that suggested improved engagement with the enhanced white space format. This format was associated with slightly longer response times, fewer re-visits between items and some improved attention to test item detail.

We argue that a nuanced approach to white space is required that recognises the space within items, between items and outside items. Increased white space between and outside items offers respondents more space to write. White space also appears to offer space to think, as frequent fixations in areas of white space in the enhanced spacing format suggest a strategy of gaze aversion to reduce cognitive load.

Some caution is required in the interpretation of our results. First, decisions about the use of white space must be balanced by wider considerations about best practice in test paper design (e.g. positioning of test items over two pages). There are also cost and environmental considerations.

It is possible that the act of asking students to compare two papers encouraged them to draw out differences, and possibly exaggerate them. Indeed, the majority of students reported that they did not notice differences in layout until probed by the interviewer. Research by Markman and Gentner (1996) suggests that when asked to make comparisons, people find it easier to list the differences for pairs of items that are similar rather than dissimilar. Our research prioritised ecological validity by using eye-tracking glasses to observe responses to paper-based examination papers. It would be useful to consider and investigate the relevance of these findings for the design and layout of computer-based assessments.

It is also worth noting that the students were not in a high-pressure, high-stakes situation, unlike when sitting their GCSE exams. It is likely that any stress-inducing effects that the white space had on the students would actually be exacerbated in a pressurised examination situation. This is arguably a strength to the work, as any stress-inducing effects observed represent a diminished effect; the reality of the issue may well be greater.

Finally, our study was small, and further validation checks with larger-scale data would be valuable to ensure that there are no significant negative consequences of changes to test paper design. For example, students with conditions such as dyslexia or ADHD were not present in our research and may be affected differently by changes in the exam paper format (Crisp, Johnson, & Novaković, 2012). It would be useful to conduct additional research in these areas.

References

Baker, W. (2001). HATS: A design procedure for routine business documents. *Business Communication Quarterly, 64*(2), 65-76.

Bax, S. (2013) The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30(4) 441–465.

Bixler, R. & D'Mello, S. K. (2016). Automatic Gaze-based User-independent Detection of Mind Wandering during Computerized Reading. *User Modeling & User-Adapted Interaction, 26*(1), 33-68.

Bolsinova, M., de Boek, P., & Tijmstra, J. (2017). Modelling conditional dependence between response time and accuracy. *Psychometrika, 82*, 1126-1148.

Bradshaw, A. C., & Johari, A. (2002). Effects of white space in learning via the web. *Journal of Educational Computing Research, 26*, 191-201.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3,* 77-101.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-130.

Crisp, V., Johnson, M. & Novaković, N. (2012): The effects of features of examination questions on the performance of students with dyslexia, *British Educational Research Journal*, 38:5, 813-839

Cubberley, C. W. (1991). Write procedures that work. *Library Journal, 116*(15), 42-45.

Davis, B. G. (1993). Quizzes, test and exams. In: B. G. Davis (Ed.) *Tools for teaching* (pp. 239-251). San Francisco, CA: Jossey-Bass Publishers.

Doherty-Sneddon, G., & Phelps, F. G. (2005) Gaze aversion: A response to cognitive or social difficulty? *Memory and Cognition, 33*, 727-733.

Ercikan, K., & Pellegrino, J. (2018). *Validation of score meaning for next generation assessments*. London, UK: Routledge.

Glenberg, A., Schroeder, J., & Robertson, D. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory and Cognition, 26*, 651-658.

Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in Raven's matrices. *Journal of Intelligence, 3*, 21-40.

Hartley, J., & Burnhill, P. (1976) Explorations in Space: A Critique of the Typography of BPS Publications. *Bulletin of the British Psychological Society, 29*, 97-107.

Hayes, N. (2000). *Doing psychological research.* Abingdon, UK: Taylor & Francis Group.

Hoener, A., Salend, S., & Kay, S.K. (1997). Creating readable handouts, worksheets, overheads, tests, review materials, study guides, and homework assignments through effective typographic design. *Teaching Exceptional Children, 29*(4), 32-35.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Hughes, A. (1989). *Testing for language teachers.* Cambridge, UK: Cambridge University Press.

Hughes, L. E., & Wilkins, A. J. (2000). Typography in children's reading schemes may be suboptimal: Evidence from measures of reading rate. *Journal of Research in Reading, 23*, 314-324.

Leube, A., Rifai, K., & Wahl, S. (2017). Sampling rate influences saccade detection in mobile eye tracking of a reading task. *Journal of Eye Movement Research*, *10*, 3.

Lonsdale, M. (2007). Does typographic design (text and question and answer sheet layout) of examination material affect performance? *Information Design Journal, 15*(2), 1-41.

Lonsdale, M., Dyson, M., & Reynolds, L. (2006). Reading in examination-type situations: the effects of text layout on performance. *Journal of Research in Reading, 29,* 433-453.

Maddox, B. (2017). Talk in Tests: A Micro-Analytic Approach to Interaction, Intimacy and Affect in Computer-Based Assessment. *Measurement: Interdisciplinary Research and Perspectives, 15*, 113-127.

Maddox, B., Bayliss, A. P., Fleming, P., Engelhardt, P. E., Edwards, G., & Bourgonovi, F. (2018). Observing response processes with eye tracking in international large-scale assessments: Evidence from the OECD PIAAC assessment. *European Journal of Psychology in Education, 33*(3), 543-558.

Maddox, B., & Zumbo, B. D. (2017). Observing testing situations: Validation as jazz. In B. D. Zumbo and A. M. Hubley (Eds.), *Understanding and Investigating Response Processes in Validation Research* (pp. 179-192). Springer Press.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*, 235-249.

Niefind, F. & Dimingen, O. (2016). Dissociating parafoveal preview benefit and parafovea-on-fovea effects during reading: A combined eye tracking and EEG study. *Psychophysiology, 53*, 1784–1798.

Oranje, A., Gorin, J., Jia, Y., & Kerr, D. (2018) Collecting, analysing, and interpreting response time, eye tracking, and log data. In K. Ercikan & J. Pellegrino (Eds.), *Validation of score meaning for next generation assessments* (pp. 39-51). London, UK: Routledge.

Quirkos (Version 1.4.0) [Computer software]. (2017). Retrieved on 31 May 2019 from http://www.quirkos.com

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506.

Rotter, K. (2006). Creating instructional materials for all pupils: Try COLA. *Intervention in School and Clinic, 41*, 273-282.

Scott, D. L., & Webb, L. F. (1979). The effects of spacing and handwritten or typed tests upon achievement of students in algebra and geometry classes. *School Science and Mathematics, 79*, 663-669.

Seli, P., Risko, E., & Smilek, D. (2016). On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychological Science, 27*, 685-691.

Strukelj, A., & Niehorster, D. C. (2018). One page of text: Eye movements during regular and thorough reading, skimming, and spell checking. *Journal of Eye Movement Research, 11*(1), 1-22.

Varao-Sousa, T., Grayden, J.F., & Kingstone, A. (2017). Re-Reading After Mind Wandering. *Canadian Journal of Experimental Psychology, 71*, 203–211.

Weir, C.J. (1993). *Understanding and developing language tests.* New York, NY: Prentice Hall Europe.

Wilson, T., Pfister, F., & Fleury, B. (1981). *The design of printed instructional materials: Research on illustrations and typography.* Syracuse, NY: ERIC Clearinghouse on Information Resources.

Zimmerman, B. B., Sudweeks, R. R., Shelley, M. F., & Wood, B. (1990). *How to prepare better tests: guidelines for university faculty.* Provo, UT: Brigham Young University.

Zumbo, B. D. (2017). Trending away from routine procedures, toward and ecologically informed in vivo view of validation practices. *Measurement: Interdisciplinary research and perspectives, 15*, 137-139.

Zumbo, B. D., & Hubley, A.M. (2017). *Understanding and Investigating Response Processes in Validation Research*. Springer Press.

Appendix

Figure A1 **Response times by score**

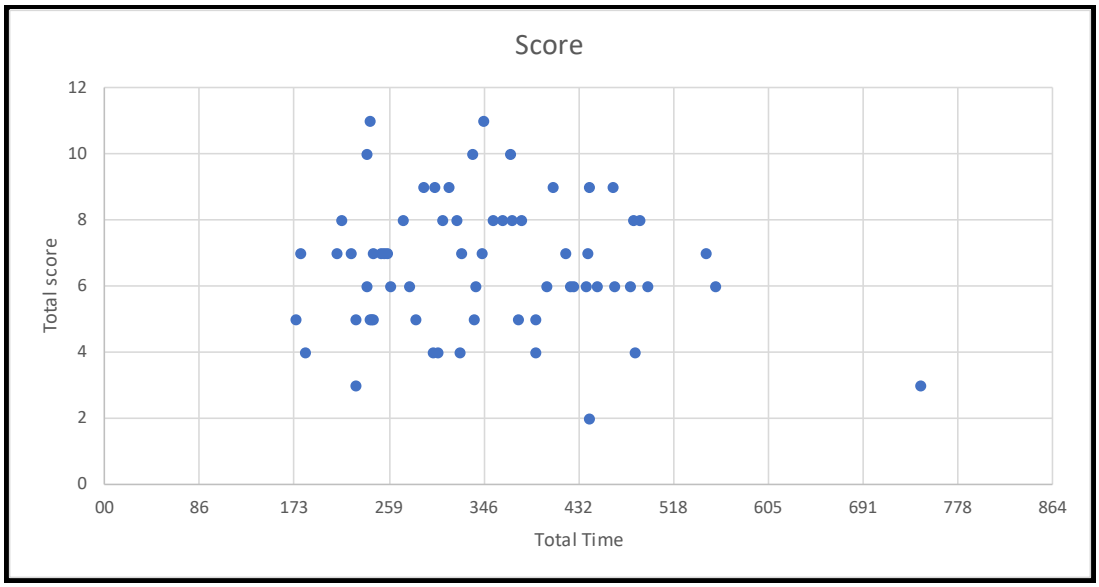


Table A1 **AOI and white space re-visits and fixation counts (Paper 1: Items 1.1–1.5)**

AOI	Type	Revisits	Fix. count
1	Restricted	3.8	24
	Enhanced	3.2	18.5
2	Restricted	6.5	56.4
	Enhanced	4.3	41.3
3	Restricted	7.9	69
	Enhanced	2.1	49.4
4	Restricted	9.8	66.7
	Enhanced	5.3	28.2
5	Restricted	7.1	77.5
	Enhanced	4	45.5
WS	Restricted	1.8	2.8
	Enhanced	6.6	13.1