Running head: Working memory in L2 input processing

The role of working memory in processing L2 input: Insights from eye-tracking

Bimali Indrarathne and Judit Kormos

Lancaster University

*We are grateful for the support of the Economic and Social Research Council, United Kingdom Grant No ES/P00962X/1. All instruments used in this study are deposited in the IRIS database (https://www.iris-database.org).

Address for correspondence:
Professor Judit Kormos
Lancaster University
Department of Linguistics and English Language
Lancaster, United Kingdom
LA1 4YL
j.kormos@lancaster.ac.uk

Attention, working memory capacity, central executive functions, input processing, eye-tracking

Abstract

Our study investigated how attention paid to a target syntactic construction *causative had* is related the storage capacity and attention regulation function of working memory (WM) and how these WM abilities moderate the change of knowledge of the target construction in different input conditions. 80 Sri Lankan learners of English were exposed to examples of the target construction in explicit and implicit learning conditions and their eye movements were tracked as they read the input. Correlational and multiple regression analyses indicated a very strong relationship between WM abilities and gains in the knowledge of the target construction. WM scores were closely associated with gains in receptive knowledge in all input conditions, but they had a weaker link to the improvement of productive knowledge in the implicit learning conditions. The amount of attention paid to input was also strongly related to WM abilities.

Introduction

Working memory (WM) ability is a key cognitive individual characteristic that can potentially influence second language (L2) learning outcomes (for a review see Juffs & Harrington, 2011). WM assists in the comprehension of L2 input, regulation of learners' attention while processing input, and encoding this perceived input into long-term memory. WM is part of what, in educational psychology, are called aptitude complexes (Snow, 1992), in other words, individual differences that dynamically interact with the situation in which learning takes place. Although the role of language learning aptitude under various learning conditions has been investigated in previous studies (e.g. De Graaff, 1997; Erlam, 2005; Robinson, 2005; Yilmaz & Granena, 2015), little is known about how the storage capacity and attention regulation functions of WM jointly influence the acquisition of previously unknown syntactic constructions in explicit and implicit learning conditions. Findings in the field of second language acquisition research (SLA) and cognitive psychology are also inconclusive with regard to the interaction of WM abilities with instructional treatment types. The novelty of our research is that it extends the use of eye-tracking methodology in SLA (e.g. Issa, Morgan-Short, Villegas & Raney, 2015; Godfroid & Uggen, 2013; Godfroid & Winke, 2015) to the analysis of how WM abilities are related to attentional processing under various instructional conditions. The main aim of our research is to examine what role the combined storage and attention regulation functions of WM play in L2 learners' attentional processing of a target grammatical construction and in accounting for learning gains in explicit and implicit instructional conditions.

In our research, L2 learners of English were exposed to the syntactic construction *causative had* (e.g. He *had* the house *painted*) in four different conditions: input flood, whereby the frequency of the target item is increased in the input; textual enhancement, when the target construction is highlighted in the text; an instruction to pay attention to the highlighted grammatical construction; and an explicit metalinguistic explanation of the highlighted target language construction with instruction to pay attention to the highlighted grammatical construction. According to Norris and Ortega (2000, p. 437), the instructional conditions that have “neither rule presentation nor directions to attend to particular forms that were part of a treatment” can be treated as implicit instruction. Thus, in this experiment, the first two can be regarded as implicit instructional conditions, and the second two as explicit conditions (Spada & Tomita, 2010). Eye-tracking was used to measure attentional processing and a novel feature of our study was that, in addition to a test of phonological short-term memory (PSTM) storage capacity, three instruments that assess individual differences in attention regulation were also administered.

Review of literature

There are two important cognitive processes associated with L2 learning that are relevant from the perspective of our study: attention and consciousness. It is widely accepted that attention paid to input is vital for L2 development (Doughty, 2001; Schmidt, 2010; Schmidt, 1990; Robinson, Mackey, Gass & Schmidt, 2012). Although there seems to be no consensus in cognitive psychology or in the SLA literature on what attention is (Allport, 1988; Hulstijn, 2015; Shinn-Cunningham, 2008; Wolfe & Horowitz; 2004), two characteristics of attentional processing seem to be key to all definitions, namely that attention is subject to intentional control and is selective (e.g. James, 1890; McFarland, 2006; Shiffrin, 1988; Smith & Kosslyn, 2006; Styles, 2006).

Consciousness, which is identified as the understanding of one’s experiences (Max Velmans, 2009; Nagel, 1974), is closely related to attention. In our paper we adapt Koch and

Tsuchiya's (2006) and Lamme's (2003) positions on attention and consciousness. Koch and Tsuchiya (2006) argue that attention with or without consciousness is possible. However, attention with consciousness is necessary so that stimuli attended to can be registered in WM, and thus people can distinguish between stimuli and provide a full report of them. Lamme's (2003) model also maintains a similar position. The model explains that input should be attended consciously if it is to enter WM (for a more detailed discussion of attention, awareness and noticing see Author 1 and 2, 2016). Chun, Golomb, and Turk-Browne (2011) distinguish external attention, which operates on stimuli perceived externally through one's senses, and internal attention, which processes information already contained in long-term memory. They argue that WM in general, and central executive (CE) functions in particular, play a key role at the interface of internal and external attentional processes. Therefore, an investigation of the association between the functioning of WM and attentional processing can yield new insights into how L2 learners attend to and process input while learning an additional language.

One of the most frequently employed WM models by Baddeley (for a recent review see Baddeley, 2015) contains three main components: a CE and two slave systems called the phonological loop and the visuospatial sketchpad. The CE acts as the general attentional controller, the phonological loop is responsible for processing information related to speech, and the visuospatial sketchpad handles information related to visuospatial imagery. In this model, the CE is considered to be a processor of all information that is not processed by the phonological loop and the visuospatial sketchpad. Baddeley (2003, 2015) explains that the CE is responsible for mechanisms such as coordinating the subsidiary memory systems, switching attention, controlling encoding, retrieval strategies, and manipulating information in the phonological loop and the visuospatial sketchpad. Some other functions are also assumed to be controlled by the CE, for example, switching attention between multiple tasks (Rogers & Monsell, 1995; Monsell, 2005), the inhibition of irrelevant information (Roberts, Hager, &

Heron, 1994), monitoring and updating the content of WM (Van der Linden, Bredart, & Beerten, 1994), temporal tagging and contextual coding of incoming information (Jonides & Smith, 1997) and planning and sequencing intended actions (Ward & Allport, 1997).

An important question relating to CE functions is to what extent its three main operations of task switching, inhibition control, and updating are inter-related with each other, and with the storage capacity of WM. According to Miyaki et al's (2000) *unity in diversity model*, these three CE functions are separable but they share a common underlying function. In a recent revision of their model, they argue that this common underlying function is "one's ability to actively maintain task goals and goal-related information and use this information to effectively bias lower level processing" (Miyaki & Friedman, 2012, p. 11). Their new research data suggest that instead of the previously assumed three separate functions of task switching, inhibition control, and updating, the diversity of the model lies in two main functions: an updating specific and a shifting specific function. They hypothesize that the updating function is responsible for the efficient manipulation and retrieval of information, whereas the shifting function reflects the flexibility with which individuals make transitions between tasks.

Although PSTM is conceptualized as being primarily responsible for storing verbal information in Baddeley's (2003, 2015) model, research evidence suggests that there is a substantial overlap between span tasks assessing PSTM and updating tasks (e.g., St Clair-Thompson & Gathercole, 2006). One important conclusion from this line of research is that in order to understand the complex role of WM abilities in learning processes and ultimate attainment, it is necessary to apply a variety of tasks that assess different processes within WM.

Based on Baddeley and Hitch's model (1974 and its extensions), Gathercole and Baddeley (1993) identify key areas of language learning, such as vocabulary acquisition, reading development, grammar learning, and speech production, which are influenced by the functioning of WM. They hypothesize that PSTM is the main component of WM that is associated with processes of language learning. It is, however, also important to study the role of CE functions because they can assist in regulating attention paid to relevant linguistic features. In addition, they play an important role in maintaining chunks of language in memory for further processing and in inhibiting irrelevant stimuli and automatic response patterns (Author 2, 2013). The inhibition function of the CE has been found to be predictive of L1 reading comprehension ability (Cain, 2006; Cain & Bignell, 2014) and associated with L2 comprehension and production processes (Abutalebi & Green, 2007). The potentially significant role of CE functions in predicting language learning outcomes and influencing L2 learning processes has led to recent reconceptualizations of the construct of language learning aptitude and to the inclusion of tests of CE functioning in assessments of aptitude (Linck et al., 2013).

Drawing on the findings of studies on the language learning difficulties faced by learners with dyslexia, who are generally characterized by a smaller PSTM capacity (Jeffries & Everatt, 2004), Ellis (1996) and Ellis and Sinclair (1996) argue that PSTM, as well as more general WM abilities, play an important role in the processes of acquiring L2 grammatical knowledge and skills. Individuals with better WM abilities are assumed to be more efficient in decoding and attending to various features in the input (Mackey, Philp, Fuji, & Tatsumi, 2002; Sunderman & Kroll, 2009). WM abilities are also implicated in maintaining linguistic information in the PSTM and in the "identification, selection, and correlation of relevant features both in the input and in long-term memory" (Martin & Ellis, 2012, p. 406).

The role of WM in grammar learning has been investigated in a large number of studies (e.g. Ahmadian, 2015; Baralt, 2015; Ellis & Sinclair; 1996; Grey, Williams & Rebuschat, 2015; Li, 2015; Martin & Ellis, 2012; O'Brien, Segalowitz, Collentine & Freed, 2006; Révész, 2012; Sagarra & Abbuhl, 2013; Santamaria & Sunderman, 2015, Tagarelli, Borges Mota & Rebuschat, 2015; Williams & Lovatt, 2003). Conflicting findings have emerged with regard to its importance in implicit learning, that is, in “the process by which knowledge about the rule governed complexities of the stimulus environment is acquired independently of conscious attempts to do so” (Reber, 1989, p. 219) and in explicit learning, which is the conscious process of gaining knowledge. Studies are also inconclusive with regard to the question whether higher WM abilities enhance the development of explicit and implicit knowledge in implicit and explicit instructional contexts.[1] Existing research findings unequivocally indicate that WM assists L2 learners in acquiring explicit knowledge, that is “facts that speakers of a language have learned” (R. Ellis, 2006, p. 95) under explicit learning conditions (Robinson, 2005; Santamaria & Sunderman, 2015). Tagarelli et al. (2015), however, found that WM capacity was not associated with the development of implicit knowledge, that is, “abstract, unconscious and rule-like representations” (N. Ellis, 2007, p. 19), in an explicit learning condition. WM was traditionally assumed to be influential in explicit learning, but not in implicit and incidental learning processes (e.g. Reber, Walkenfeld & Hernstadt, 1991), which are the by-products of other cognitive operations and less prone to the influence of individual differences (Unsworth & Engle, 2005). Nonetheless, recent work in the field of cognitive psychology by Hassin, Bargh, Engell and McCulloch (2009) and Soto and Silvanto (2014) indicates that WM operations might be involved in implicit learning processes. In implicit learning conditions high working memory capacity can assist in keeping larger chunks of information active for further processing and can extend the scope of attention (Martini, Sachse, Furtner, & Gaschler, 2015).

Findings with regard to the role of WM abilities in the implicit learning of L2 grammar have also been mixed. Learning gains in explicit knowledge in implicit learning contexts were significantly associated with higher WM abilities in most studies (Ellis & Sinclair, 1996;

---

1. Implicit instruction is “directed at enabling learners to infer rules without awareness” (R. Ellis, 2009, p. 16) and in explicit instruction “learners are encouraged to develop metalinguistic awareness of the rule” either inductively or deductively (ibid, p. 17).

Martin & Ellis, 2012; Williams & Lovatt 2003) except for Grey et al.'s (2015) research. A significant link between WM abilities and implicit knowledge gain in an implicit learning condition was also found by Ellis and Sinclair's (1996) study. In contrast, Robinson's (2005), Tagarelli et al.'s (2011) and Grey et al.'s (2015) research revealed no relationship between WM abilities and the development of implicit knowledge in implicit learning conditions.

Given the contradictory findings in the field of SLA research with regard to the importance of WM abilities in grammar learning under different conditions, and the dominance of previous research that only assessed the storage function of WM, it is important to examine what role WM abilities, including both PSTM storage capacity and CE functions, play in L2 learners' attentional processing of a novel grammatical construction. All previous studies exposed participants to either an artificial language or a language unknown to them and presented grammatical constructions out of context. Therefore, little is known about how WM abilities affect grammatical development in a language that the participants have prior knowledge of and how these cognitive factors influence processing of a novel grammatical construction that is embedded in a reading text.

The chosen syntactic construction was *causative had*. The choice of this construction was motivated by the assumption that the selected pre-intermediate learner sample of Sri Lankan students would have no or very little pre-existing knowledge of it as this construction is usually taught at higher proficiency levels in English language courses. Another important consideration was that the construction should form an easily identifiable area of interest for eye-tracking research. Furthermore, we also aimed to select a construction whose meaning can be inferred from the context, and for which there is a one-to-one form to meaning mapping. Moreover, the causative construction in Sinhala, the first language of the participants of the study, is different to that in English. Sinhalese is a SOV language, and the causative is marked by bound suffixes.

In our study, three different tests of the CE and a test assessing the storage capacity of PSTM were applied. With the administration of four different assessment tools, we aimed to overcome a shortcoming of many previous studies that used instruments that only measure the storage capacity of PSTM. The inclusion of tests of the CE also allowed us to examine its additional role in the regulation of attentional processing in combination with the PSTM.

Our research questions were:

RQ 1: How is the functioning of WM related to the change in knowledge of the target grammatical construction 'causative had' in different input conditions?

RQ 2: How is the functioning of the WM related to the attention paid to target items?

Method

Context

The data for this study were collected in Sri Lanka at a state university where the student population comes from Sinhala or Tamil first language speaking backgrounds. Most of these students belonged to the first language Sinhala community, which receives their primary and secondary education at Sinhala medium schools. They learn English as a subject from Grade 1 to university entrance, i.e., for approximately 13 years. The medium of instruction at this university is English for both undergraduate and postgraduate courses and all undergraduate students have to take English language as a credit-bearing course throughout the degree programme.

Participants

The original sample of participants in this study included 103 first year undergraduates from the Bachelor of Commerce degree programme, and their ages ranged from 19 to 21 years. Based on university entrance test results, the students had around B1/low B2 level of language proficiency on the Common European Framework of Reference (Council of Europe, 2001). By the time data collection was completed, the participants had been learning English at the university for five months. The 29 female and 71 male participants all belonged to the first language Sinhala community. English was the only additional language these participants had learned. None of the participants of the study were bilinguals according to the questionnaire data.

Three students were excluded from the study because they demonstrated existing knowledge of the target condition in the sentence reconstruction (SR) task of the pre-test, i.e. they scored 2 or above in this test (see below). Thus, as previously stated, 100 participants were invited to participate in further stages of the research. Among them 97 scored 0 in this component of the pre-test. The other three students scored 1 in the item with an irregular past participle (*put*). Thus, it was assumed that they did not have knowledge of the target construction, and they gave a correct answer to this item by chance. Eighty students were randomly assigned to the experimental conditions and 20 learners participated as controls. The participants who scored 1 in the SR task were in three different input conditions. As we explain below, high-quality eye-tracking data were available for 45 out of the 80 experimental group participants. The control group did not participate in the eye-tracking phase of the study.

Materials

Input texts. Three short stories were applied as input texts, which we had written specifically for the purposes of this study (for an example see the supplementary file). After preparing the text of the stories, Vocabprofile (Heatley, Nation, & Coxhead, 2002) was used to determine the lexical complexity of each text. More than 90% of the words in all texts were K1 words (most frequent 1,000 words in English) and K2 words (the second most frequent 1,000 words in English). For example, Text 1 had 93.56% K1 and K2 words, Text 2 had 91.74% of K1 and K2 words, and Text 3 had 91.84% of K1 and K2 words. The rest of the words, which Vocabprofile identified as 'off list words', were proper nouns. This procedure assured that the three texts were within the lexical competence of the participants and contained very frequent words. This was important because the participants would have otherwise spent more time gazing at unknown vocabulary in the text which would have given inaccurate eye-tracking data of the amount of attention paid to the target examples. The Flesh-Kincaid Grade Level value was 4 for all three texts.

The target construction was *causative had* (had + article + noun + main verb). Seven examples of the target construction were included in each input story, with a total of 21 items in the three texts. In order to ensure that the verbs used in the examples of the target construction belonged to the most frequent verbs used in the particular construction, all target items were taken from the British National Corpus (BNC). It was also necessary to maintain the storyline; therefore, some alterations were made to the nouns (e.g. had the *tools* delivered instead of had the *letters* delivered). We also controlled the target items for syllable length so that all of them contained four to eight syllables (see supplementary materials for BNC frequencies and the numbers of syllables in the target examples).

Four reading comprehension questions were included at the end of each text with the aim of giving the participants a purpose for reading. Two of the four questions measured if the participants understood the meaning of the target construction (e.g., Who stopped the bus?) and the other two items assessed their general understanding of the text.

Pre-test and post-test. A sentence reconstruction (SR) task and a grammaticality judgment (GJ) task were used as pre- and post-tests. These tests were selected so as to be able to assess the comprehension and production of the target construction. In the SR task the participants were asked to construct a sentence with a similar meaning to a previously given sentence. The first words of the sentence to be produced were given as a cue.

A sample SR item:

Sara got someone to print invitation cards for her party.
Sara had ....

In the GJ task, the participants were asked to judge whether the sentences were grammatically accurate and tick the relevant column.

A sample GJ item:

My dad had his lunch delivered to his office yesterday. Correct/Incorrect

We used 20 SR items of which six were target items and 40 GJ items of which ten were target items, while others functioned as distracter items in both pre- and post-tests. The ungrammatical items included two items where the –ed suffix was omitted, three word order errors and one instance of the use of the wrong suffix, –ing. The length of sentences was also controlled: all items except two GJ items contained ten words each and the remaining two items consisted of eleven words. The SR task was administered in a written format with a time limit and the participants were expected to use the target construction in their reformulated sentences. The GJ task was administered aurally with the help of a recorded presentation of the 40 items. One of the authors, whose first language was the same as that of the participants, recorded the sentences with a five-second interval between each of them. We wanted to ensure that the participants understand the sentences that they heard and that they did not have comprehension difficulties because they were not familiar with the accent of the speaker.

In the GJ test, the participants were expected to decide on the grammatical accuracy of the sentences that they heard and mark it on an answer sheet. This allowed us to maintain some degree of time pressure in this task. The nouns in subject and object positions used in the pre-test sentences were changed in the post-test sentences to minimise the possibility that the participants would answer the items based on their memory of the pre-test items. The BNC was consulted to construct sentences for the pre- and post-tests, but we made slight alterations to nouns in the original BNC sentences, and added some phrases, in order to maintain the equal length of sentences.

The reliability of the post-test SR task as measured by Cronbach's alpha was 0.681 and that of the post-test GJ task .447. The lower value of the GJ task might have been due to the lower number of items (Lance, Butts & Michels, 2006). It is also important to note that the Cronbach alpha measure assesses internal consistency across test items and might underestimate reliability if the instrument is not unidimensional or uses dichotomous items. Our test items had only two values (correct vs. incorrect) and included both grammatical and ungrammatical items, and within the ungrammatical items the grammatical errors were also different. The Cronbach alpha for the gain scores, which were used in the correlational and regression analyses, were above .6 (SR task Cronbach alpha = 0.692; GJ task Cronbach alpha = .650). Given the low number of items, these values can be considered acceptable. The mean for the post GJ task ($M = 5.80$, $SD = 2.02$) was also significantly above the 50% guessing range ($t\,(96) = -3.79$, $p < .001$).

Working memory tests. Four different cognitive tests were used in this study, which were selected based on previous research in the field of cognitive psychology. The forward digit-span test has been used by Briscoe and Rankin (2009), Gathercole, Willis, Baddeley and Emslie (1994), Pickering and Gathercole (2001) and Shahabi, Abad and Colom (2014) to assess the storage capacity of the PSTM. Updating is often measured by Keep Track tasks

(Shahabi, Abad, & Colom, 2014; Tamnes et al., 2013; Wilhelm, Hildebrandt, & Oberauer, 2013; Zheng et al., 2012) and Plus Minus tasks have been applied to assess attention shifting capacity (Hull, Martin, Beier, Lane, & Hamilton, 2008; Puric & Pavlovic, 2012; St Clair-Thompson & Gathercole, 2006; Tamnes et al., 2013). The Stroop task has been used widely in psychology as a tool to measure inhibition (Shahabi et al., 2014; Zheng et al., 2012).

The forward Digit-Span test used in this study consisted of nine spans starting with three digits and going up to eleven digits. Each span consisted of two sets. Before taking the test, the participants completed a set of practice tasks, which consisted of two three-digit items. In the test the longest span that the participants could remember without a mistake was taken as their actual digit span.

Based on Miyake et al.'s (2000) description, a Keep Track task was used in this study to assess the updating function of the CE (for an example of this task see the supplementary file). In completing this task, participants have to monitor incoming information, delete irrelevant information, and update it with new and relevant information. In this task, words from six different categories – animals, colours, countries, distances, metals, and relatives – were used. Each category contained three words (for example, animal words were cat, horse, cow). First, the participants were shown all the categories (animals, colours, countries, distances, metals, and relatives) and all the words within each category that would be used in the experiment. Then they were told that they would be presented with words from one category in a serial random order and their task was to remember the last word of the category. Then they had to write down the last words of the categories shown to them in each trial. The target category (e.g., country) was given at the bottom of the screen each time when the participants were shown the words. The total number of words that the participants had to recall was 27 in six different series (three trials with words from four categories and three trials with words from five categories) excluding two practice trials with words from three different categories. The proportion of correctly recalled words was taken as a measure of the updating function of the CE. In this test the words were in English, but they were highly frequent and well-known by the participants,.

An Internet-based Stroop task (Online version – cognitivefun.net-test2) was used to measure inhibition. In this task, the participants were shown the names of colours (e.g., red, blue, purple) and the letters on the screen also appeared in a particular colour. They had to press the first letter of the colour of the letters that they could see, not the first letter of the colour mentioned in the word (e.g., if the letters that appeared on the screen were ‘green’, but if the colour of those letters was in red, they had to press ‘r’ on the keyboard). The measurement made was of ‘interference time’, i.e., the difference between the time taken to press a key when a word was written in the colour of the letters and when the word did not match the colour of the letters. The program provided information on accuracy rate, overall reaction time, and the interference time.

Based on the description provided by Miyake et al. (2000), a Plus Minus task was devised to measure attention shifting. The participants were given three lists of two-digit numbers (30 in total in each list): in the first list they had to add 3 to each number and write the answer, in the second list they were asked to subtract 3 from each number and write down the answer, and in the third list they had to shift between adding and subtracting 3 from alternative numbers. The participants were asked to complete each list as quickly as possible and were told to pay attention to the accuracy of their answers. The time spent on completing each list was measured by a digital stopwatch, which provided data measured to 1/100 sec with 0.003% accuracy. The shifting cost between addition and subtraction was “calculated as the difference between the time to complete the alternating list and the average of the times to complete the addition and subtraction lists” (Miyake et al., 2000, p. 65).

Procedures. The study had five groups to which we randomly assigned 20 participants each. Four were experimental groups and one was the control group. On the first day of data collection, the 100 participants were given a consent form to sign and a background questionnaire to fill in. They also took the pre-test on that day. Five days later, the eye-tracking sessions started, and the participants in the four experimental groups met with the first author every other day on three occasions to read the three texts on the eye-tracker. The third eye-tracking session was immediately followed by a post-test. Three days after the post-test, they took WM tests. The control group took only the pre and post-tests and did not participate in the WM tests. In the pre- and post-tests, the participants first completed the SR and then the GJ task.

Eye-tracking procedures. As illustrated in Figure 1, the four experimental groups were provided with input in four different ways. The examples of the target construction in the three texts were textually enhanced by bolding for three of the groups, while one group read unenhanced texts. At the beginning of each eye-tracking session, the *enhanced+instructions* and the *enhanced+instructions+explanation* groups were requested to pay attention to the phrases highlighted in bold. In addition, participants in the *enhanced+instructions+explanation* group received an explanation of the form and meaning of the target construction in a PowerPoint presentation immediately before the second eye-tracking session. The presentation consisted of six slides and included four examples from Text 1. The use of the same examples as in the input text ensured that this group would not be exposed to substantially more input in addition to the metalinguistic explanation than the other experimental groups. The *enhanced only* and *unenhanced* groups were not instructed to pay attention to any specific aspect of the input texts.

Insert Figure 1 around here

A Tobii X2-60 portable eye-tracker was used for the collection of eye-tracking data. We placed the laptop on which the eye-tracker was fixed in a quiet room where individual participants read the input texts. At the beginning of the experiment, the first author explained the functions of the eye-tracker to the participants who were then allowed to go through a trial slide to ensure that they understood the procedure. A 9-point calibration was carried out at the beginning of each session. Participants sat approximately 67 centimetres away from the screen.

The eye-tracking slides were first prepared in PowerPoint using 24-point, double-spaced Calibri font. We selected this font because it is considered to be suitable for screen display (Ericson, 2013). Since we were interested in the total fixation duration on the whole construction (AOI or Area of Interest, e.g., he had the walls painted), all the words that belonged to the target construction were placed in one line. The position of the AOIs within the lines, however, could not be controlled because the targets formed part of a coherent storyline. As the targets and the texts were identical in all experimental conditions, variations in the position of the AOIs were not expected to influence the findings. Each slide contained 4–5 double-spaced lines (see Figure 2 for an illustration of a sample of an unenhanced input slide and Figure 3 for a sample of enhanced input). At the beginning of each session, the participants were informed that they were going to read a story. During the eye-tracking session, they were not allowed to move back to a previous slide; however, they could spend as much time as they required on the current slide.

Insert Figures 2 and 3 around here

Working memory testing procedures. The participants took the four WM tests individually in a quiet room with the researcher present. In the digit-span test, first they read the instructions slide in the PowerPoint presentation, and the first author of the paper also gave a verbal explanation. Then the participants completed the first two practice spans and following this they took the test. In this task, digits were presented in the PowerPoint slides (one number per slide) with 0.1 seconds between them. The participants had to look at the screen and write down the span when the instruction slide “Write” appeared after each span.

In the Keep Track task a similar method was applied using a PowerPoint presentation. The participants read the instructions slide and then listened to the instructions of the researcher. Next, they completed the two trials and the researcher showed them the answers to ensure that the participants understood the instructions. Following this, the test was administered to the participants. In the Stroop task, the participants received verbal instructions on how to perform the task. Then, they had to complete the task by pressing the correct keys on the computer keyboard for 30 trials. The Plus Minus task was completed on paper. First, the participants read the instructions and also listened to verbal instructions given by the researcher. Then they were asked to complete the first column asking them to add 3 to the numbers given. The time to complete the column (in milliseconds) was measured by a stopwatch. The same process was applied in completing the other two columns.

Data analysis. We used two eye-tracking measurements to determine the amount of attention paid to the examples of target construction. The first was Total Fixation Duration (TFD). First the mean TFD for each participant was computed. For this calculation, the TFD of each participant on each AOI was extracted from the Tobii software and a mean value calculated. For the second eye-tracking measurement, first the TFD for each participant for the whole page was extracted. Then, the expected TFD on each AOI was calculated based on the proportion of the number of syllables that the AOI had in relation to the number of syllables on the whole page where the particular AOI occurred.

$$\text{Expected fixation duration of an AOI} = \frac{\text{No. of syllables of the AOI} \times \text{TFD-whole page}}{\text{No. of syllables of the whole page}}$$

As a next step, the difference between the observed and expected TFD was calculated for each participant for each AOI. Finally, the mean value of the difference for each participant for all AOIs was computed and used in further analyses. This is presented as the difference between observed and expected fixation duration (ΔOE). For example, if a participant spent 2400 milliseconds on reading a slide that contained a text with 30 syllables, the expected fixation duration for the target construction that consisted of 10 syllables was 800 milliseconds (10/30 x 2400). If the participant's actual fixation duration was 900 milliseconds, then the ΔOE value was 100 milliseconds (900-800).

Syllables were used because Trueswell, Tanenhaus, and Garnsey (1994) argue that letter-based standardizations are inaccurate. Furthermore, in English, there is no one-to-one correspondence between letters and sounds and often letter combinations rather than individual letters denote phonemes. This results in the fact that beyond the initial stages of learning to read, it is not letters but larger grain-sized units such as onsets and rimes and ultimately syllables that constitute the basic units of word-level decoding in English (Ziegler & Goswami, 2005).

Unfortunately, on each testing occasion, approximately 23-25 % of participants' data had to be excluded because their eye-fixations went beyond the screen for a considerable amount of time. This is due to the limitations of the portable eye-tracker used to collect data (Tobii X2-60). Use of a desktop eye-tracker would have prevented the data loss. As we needed information for all three text-reading sessions, only those participants whose data were of appropriate quality for all three texts were included in the eye-tracking data analysis for Research Question 2 (see Table 1). This resulted in the loss of data for 35 participants out of the 80 whose eye-tracking measures were taken in the experimental conditions.

Insert Table 1 around here

Scoring of the pre- and post-tests and the working memory tests. When scoring both SR and GJ items, spelling mistakes were ignored and half marks were not awarded. Each correct answer received one mark. The gain scores in both SR and GJ tasks were computed separately as the difference between the total post-test score and the total pre-test score.

In the Digit-Span test, the longest span participants could remember without a mistake was taken as their PSTM span. The total number of words that they had to recall in the Keep Track task was 27. The proportion of correctly recalled words was taken as the measure of updating. In the Stroop test, the interference time was used as a measure of inhibition. The average interference time in milliseconds for each participant as calculated and displayed by the website itself was taken as the score for this test. In the Plus Minus test, the shifting cost was used as the measurement.

Results

First of all, before analysing the role of PSTM storage and CE functions in processing L2 input, we checked if the treatment groups were comparable in terms of their cognitive abilities. The MANOVA analysis showed that the four groups were not statistically different from each other with regard to any of the cognitive tests $F(3, 75) = 1.06$, $p = .391$ (see Table 2 for descriptive statistics). As a next step we computed correlations among the three tests of CE functions and PSTM. As can be seen in Table 3, the Digit-Span and Keep Track tasks were very highly correlated. There was still a significant and strong correlation between the Keep Track, and the Stroop tasks, and the Stroop and the Digit Span tasks, but the magnitude of these correlations was smaller. The performance on the Plus Minus task was not linked to the scores of any of the other cognitive tests. In order to test Miyaki et al's (2000) *unity in diversity* hypothesis on our data, we conducted a principal component analysis to examine the factorial structure of these four cognitive tests. The factor analysis indicated that the Digit Span, Keep Track and Stroop tasks form one factor (Eigenvalue= 2.42) and the Plus Minus task constitutes a different factor (Eigenvalue = 1.02). The Kaiser-Meyer-Olkin measure of sampling adequacy was .68, which is higher than the recommended minimum value of .50 (Pett, Lackey, & Sullivan, 2003). The Barlett's test of sphericity reached statistical significance (p < .001), supporting the factorability of the correlation matrix. The common variance shared by the Digit Span (Factor 1 loading = .939), Keep Track (Factor 1 loading = .903) and Stroop (Factor 1 loading = -.850) was 60.53%. The Plus Minus (Factor 2 loading = .999) task added 25.08% variance to the model. The results of the correlational and factor analysis lend very strong support for Miyaki and Friedman's (2012) revised unity and diversity model as they suggest that attention switching seems to be a distinct component of CE functions, whereas storage capacity, updating, and inhibition control are interdependent. Based on the commonalities among the Digit Span, Keep Track, and Stroop tasks and the results of the factor analysis, it was deemed appropriate to create a composite score using regression factor scores (Tabachnick & Fidell, 2001). An additional motivation for the use of this factor score was to avoid a Type I error by applying multiple tests with highly inter-correlated variables.

Insert Tables 2 and 3 around here

Relationship between gain scores and WM abilities

In order to answer the question of how the functioning of WM is related to the change in knowledge of the target grammatical construction 'causative had' in different input conditions, we first analysed learning effects, which are reported in detail in Author 1 and Author 2 (2016). For this analysis we applied mixed between-within subjects analysis to assess the impact of the four different treatment conditions (*enhanced+instructions, enhanced+instruction+explanation, enhanced, unenhanced, control*) across two time periods (pre- and post-test) (see descriptive statistics in Table 4). For the SR task, the analysis showed a significant effect of the treatment condition with a large effect size, $F(4, 95) = 8.22$, $p < .001$ partial eta squared = .25, a significant effect of time Wilks' Lambda = .57 $F(4, 95) = 66.539$, $p < .001$ partial eta squared = .425, and a significant interaction between treatment conditions and time also with a large effect size, Wilks Lambda = .71, $F(4, 95) = 9.62$, $p < .001$ partial eta squared = .288. A follow-up ANOVA with the gain score as the dependent variable and the post-hoc Bonferroni analysis indicated a significant mean difference in the SR gain scores between the control group and the enhanced+ instructions group ($p = .018$) and between the control group and the enhanced+ instructions+ explanation group ($p < .001$). A significant mean difference was also found between the unenhanced group and the enhanced+ instructions group ($p = .012$) and the enhanced+ instructions+ explanation group ($p < .001$). Moreover, the enhanced+ instructions+ explanation group showed a significant improvement ($p = .001$) compared to the enhanced only group.

The analyses for the GJ task indicated a significant effect of time Wilks' Lambda = 0.57 $F(4, 95) = 66.042$, $p < .001$ partial eta squared = .423, but no significant treatment condition effect, $F(4, 95) = 1.28$, $p = .281$, partial eta squared = .051. However, a significant interaction between treatment conditions and time was found, Wilks' Lambda = 0.81, $F(4, 95) = 5.48$, $p = .001$, partial eta squared = .188. The comparison of the gain scores of the groups by means of ANOVA and post-hoc Bonferroni tests showed that *enhanced + instructions* group (*p* =.005), and the *enhanced+ instructions+ explanation group* (*p* = .001) and *the enhanced* group (*p* = .036) achieved significantly higher gains than the control group (for further details see Author 1 & 2, in 2016).

Insert Figure 4 and Table 4 around here

Next we conducted a correlational analysis with the composite WM score, the Plus Minus task, and the gains in the SR and GJ tasks using Spearman rank-order correlations. Spearman rank-order correlations were selected because for the group-level analyses the sample size did not warrant the application of Pearson correlation analyses. As Table 5 indicates, the correlations between the composite WM score and the gains in both tasks were statistically significant and were particularly strong in the *enhanced+instructions+explanation* group. The weakest relationships emerged in the *unenhanced* group (see Figure 4). The Plus Minus task only correlated significantly with the gain in the SR task in the *enhanced+instructions* group.

Insert Table 5 around here

To assess whether there was a differential effect of WM abilities on gains across the four experimental conditions, we also ran multiregression models with the experimental condition and the WM composite score as the independent variables. The dependent variable for the first model was the SR gain score and for the second model the GJ test gain score. The *unenhanced* group was used as the reference category in both models. This analysis complements the above described correlational analyses as it allows us to directly compare the strength of the association between gain scores and WM abilities across experimental conditions. In the case of gains in the SR task, the multiple regression model showed a significant interaction effect between the treatment type and the composite WM score (Wald $\chi^2 = 23.089$, $p < .001$) suggesting that WM abilities have a differential effect on SR gain scores in the different experimental groups. The analysis revealed that the relationship between the SR gain score and the composite WM score was different in the *enhanced+instructions* (β=1.105, $p < .001$) and the *enhanced+instructions+explanation* group (β=.973, p <.001) when compared to the relationship in the *unenhanced* group. This indicates that WM abilities played a stronger role in performance gains in the SR task in the *enhanced+instructions* and *enhanced+instructions+explanation* groups than in the *unenhanced* group. In the GJ task, however, we found no significant interaction between the WM composite score and the differential instructional conditions score (Wald $\chi^2 = 1.114$, $p = .766$) (see Figure 5). This indicates that WM abilities had a similar effect on GJ test gain scores across the experimental conditions.

Insert Figure 5 here

Relationships between eye-tracking measures and WM abilities

With our second research question, we aimed to investigate the nature of the relationship between WM capacity and the attention paid to the target language item in different input conditions. For this purpose, we first conducted correlational analyses using Spearman rank-order correlations (see Table 7 for the correlations and Table 6 for descriptive statistics). When the whole sample was considered, the correlations between the two eye-tracking measures and the composite WM scores were significant, but weak. The Plus Minus task did not correlate significantly with either the TFD or ΔOE values. In the *enhanced+instructions+explanation* group the strength of the relationship between the fixation durations and the composite WM score was remarkably high. The ΔOE value also showed a strong correlation with the composite WM score in the *enhanced+instructions* and *enhanced only* groups.

Insert Tables 6 and 7 around here

For an examination of the interaction between WM abilities and experimental conditions, we ran a multiregression model with the eye-tracking measures as outcome variables, and the experimental condition and the WM composite score as the independent variables. Similar to the analysis of gain scores, this analysis yielded insights into whether the strength of association between the composite WM scores and eye-tracking measures differs in the four treatment conditions. The *unenhanced* group was used as the reference category. In the case of composite TFD measure as the dependent variable, the multiple regression model showed a significant interaction effect between the treatment type and the composite WM score, suggesting that the strength of association between the TFD measure and the composite WM score varied significantly across the experimental groups (Wald χ^2 = 34.49, p < .001) (see Figure 6). The analysis also revealed that the relationship between the TFD and the composite WM score in the *enhanced+instructions* (β =. 274, p =.042) and *enhanced+instructions+explanation* groups (β = .723, p <.001) was statistically different from the relationship observed in the *unenhanced* group. As can be seen in Figure 7, similar findings were obtained for the composite ΔOE score as the dependent variable (Wald χ^2 = 29.178, p < .001). The relationship between the composite TFD and the composite WM score in the *enhanced+instructions* (β =.395, p =.001) and *enhanced+instructions+explanation* groups (β = .608, p < .001) also differed from the relationship observed in the *unenhanced* group.

Insert Figures 6 and 7 around here

Discussion and conclusions

Role of working memory abilities on learning gains under different instructional conditions

In our first research question we asked what role individual differences in the functioning of WM play in participants' learning gains in a previously unknown syntactic construction in various instructional conditions. For the whole sample the correlational and multiple regression analyses suggest a very strong relationship between the composite score of the examined WM abilities and gains in both the SR and GJ tasks. This indicates that when the scores of abilities relating to storage capacity, updating, and inhibition are combined, they account for a substantial variance in gains in these tests. The results relating to the Plus Minus task reveal that the ability to switch between tasks does not play an influential role in the acquisition of the targeted grammatical construction. The only significant correlation for this cognitive test emerged in the *enhanced+instructions* group, and it indicated an inverse relationship between task switching and the SR gain score. This suggests that flexibility to switch from one task to the other seems to be advantageous with regard to the acquisition of novel grammatical constructions when students are explicitly told to pay attention to specific features of the input.

The use of tests CE functions to assess L2 learners' aptitude complexes was first recommended by Linck, Osthus, Koeth, and Bunting (2013), but in their study measures of the CE had little predictive power for ultimate attainment in L2 reading and listening skills. To our knowledge, our study is the first one to show that a combination of WM tests that assess the storage as well as the updating and inhabitation functions can jointly explain a considerable amount of variation in the development of syntactic knowledge through exposure to exemplars in meaningful written texts. This points to the significance of the ability to regulate attentional resources efficiently in learning from written L2 input. Higher WM abilities have been associated with better L2 reading skills in previous research (e.g., Alptekin & Ercetin, 2009; Harrington & Sawyer, 1992; Miyake & Freedman, 1998). Therefore, we can hypothesize that learners with a strong ability to hold and update verbal information in short-term memory and to inhibit irrelevant stimuli have more attentional resources available for processing grammatical information in a reading text such as the one used in our study. In addition, the increased storage and processing functions of WM allow learners to manipulate larger chunks of language more efficiently. This assists pattern recognition as well as encoding grammatical knowledge in long-term memory (Martin & Ellis, 2012; Williams & Lovatt, 2003).

The results of the correlational and multiple regression analyses conducted with the whole sample also suggest that the composite WM scores account for a slightly higher percentage of variance in the GJ task than in the SR task. However, when the correlation co-efficients for the composite WM score (SR gain r = .614 and GJ gain r = .658) are compared, no statistically significant differences emerge between them (z = -.46, p = .64). The GJ task, which was presented to the learners in auditory mode, was carried out under some degree of time pressure and assessed the receptive knowledge of the grammatical construction. In contrast, the SR task required the participants to use their knowledge of the target construction productively, and was not performed under time pressure. Therefore, our results indicate that WM abilities played an important role in the development of grammatical knowledge regardless of whether participants had to apply this knowledge productively or receptively.

Our results also show that the combined storage, updating, and inhibition functions of WM contribute significantly to learning gains in all four learning conditions. The multiregression analysis revealed that the effect of the composite WM score was similar across groups in the case of the GJ task. Therefore, learners with high WM storage capacity and good abilities to regulate their attentional resources seem to be at an advantage in acquiring receptive grammatical knowledge regardless of the type of instruction they received. Our results are similar to those of Ellis and Sinclair's (1996) with regard to the facilitative role of WM abilities in implicit learning conditions. In the SR task, however, WM abilities seemed to play a different role in the different experimental groups. In the *unenhanced* condition, which can be considered to be an implicit learning context, the contribution of the composite WM scores to improvement in SR task gains was significantly smaller than in the *enhanced+instructions* and *enhanced+instructions+explanation* conditions, which were the two highly explicit learning contexts. We might hypothesize that perhaps the productive use of grammatical knowledge acquired in implicit input conditions is less influenced by WM storage and processing abilities, and other components of aptitude, such as inductive learning ability, play a more important role in these contexts (see also Skehan, 2015).

Relationship of eye-tracking measures and working memory abilities

An examination of the relationship between WM abilities and eye-tracking measures, which was the focus of our second research question, can help us to elucidate the potential mechanisms of how the storage, updating, and inhibition functions of WM can affect learning in implicit and explicit learning conditions. When we analysed the correlations between eye-fixation measures and the composite WM factor score for the whole sample, we found significant but moderately strong correlations with the ΔOE measure. The relatively weak relationship across the sample can be explained by the fact that our previous study showed a large effect of the treatment condition on eye-tracking measures, TFD, (F = 27.06, p < .001, partial eta squared = .67) and ΔOE (F= 28.42, p < .001, partial eta squared = .68) (see Authors 1 & 2, 2016) with the *enhanced only* and *unenhanced* groups demonstrating very low levels of attentional processing of the target input. Our results reveal that in the most explicit learning condition, which was the *enhanced+instructions+explanation* condition, individual differences in WM abilities are highly predictive of how much attention L2 learners pay to the target input. The ΔOE measure, which we argue to be an indicator of attentional processing in addition to what would be expected if learners were just simply reading the text for meaning, was also strongly associated with the composite WM score in the *enhanced only* and *enhanced+instructions* groups. This finding reveals that those learners who have high WM storage capacity and efficient attention regulation abilities engage in more attentional processing of the input if their awareness of the existence of a target syntactic construction in the input is experimentally manipulated. In other words, these learners might respond better to the instruction that they should pay specific attention to the target construction embedded in the text because they are more efficient in directing their attentional resources.

What is interesting in our study is that despite the apparent lack of increased attentional processing of the target language input in the *unenhanced* condition, the correlational analyses show that L2 learners with better WM abilities still made more learning gains in both receptive and productive grammatical knowledge. This suggests that WM abilities can potentially be involved in implicit L2 learning processes. Recent theorizations of WM by Hassin et al. (2009) and Soto and Silvanto (2014) question the previously held assumption that WM can only operate on consciously attended input (e.g. Baddeley, 2003; Koch & Tsuchiya, 2006). Emerging evidence in Hassin et al.'s and Soto and Silvanto's work suggests that conscious awareness might not be necessary for information to enter WM and hence WM can also be active in implicit learning. As Soto and Silvanto highlight, "The key aspect is that WM processes can be engaged without awareness, that WM can operate on non-conscious representations, and that the process through which WM contents reach awareness is subject to varying factors (i.e. attention, intention, motivation, heuristics)" (p. 524). In our study we found a relatively strong association between the composite WM score and gains in a test of receptive knowledge in the implicit learning condition in the absence of a link between attentional processing and WM abilities. This finding seems to lend additional support to the relevance of WM abilities in implicit L2 learning.

In sum, our study shows that L2 learners with better WM abilities are advantaged in learning a novel grammatical construction from written input in both explicit and implicit conditions. Combined storage, updating and inhibition functions of WM, however, were found to play a larger role in the acquisition of productive knowledge in explicit learning conditions than in implicit conditions. As regards the development of receptive knowledge, there did not seem to be a difference in the strength of association between the composite WM score and learning gains in the different treatment groups. This suggests that regardless of the type of instruction, language learners with better WM abilities improve their receptive knowledge more successfully than those with lower levels of storage capacity and updating, and inhibition functions. These findings have direct relevance for language teaching pedagogy as they indicate that this latter group of L2 learners might require extensive exposure to target syntactic constructions and additional instructional support.

Our study has a number of limitations, the most important being the fact that a substantial proportion of the learners' eye-tracking data had to be excluded and hence the sample size for which we could analyse eye-tracking measures within the specific treatment groups was relatively small. It would have been useful to administer a post-test to the learners for us to be able to assess long-term learning gains, but unfortunately for logistical reasons this was not possible. A larger battery of WM tests could also have been applied so that we could have drawn conclusions about participants' WM abilities based on more sources of information. It would also seem to be important to examine to what extent our conclusions concerning the role of WM abilities in grammar learning can be upheld with other syntactic constructions and when input is provided through an aural rather than a written mode. We controlled for unknown vocabulary in the reading texts used; however, it does not guarantee that all learners knew all the words used. Nevertheless, our study is the first one in the field that has assessed the combined role of storage, updating, and inhibition functions of WM in developing grammatical knowledge through meaningful reading input in a language the participants had prior knowledge of. It would be useful to conduct further studies that use authentic language input in a learning context that resembles naturalistic and classroom language learning more closely.

Notes

1 In this paper, the four experimental conditions are named as follows: input flood-*unenhanced*, textual enhancement-*enhanced only*, an instruction to pay attention to the highlighted grammatical construction-*enhanced+instruction*, an explicit metalinguistic explanation of the highlighted target language construction-*enhanced+instruction+explanation*

2 Implicit instruction is “directed at enabling learners to infer rules without awareness” (R. Ellis, 2009, p. 16) and in explicit instruction “learners are encouraged to develop metalinguistic awareness of the rule” either inductively or deductively (ibid, p. 17).

References

Author 1 and 2 (2016)

Author 2 (2013)

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20,* 242–275.

Ahmadian, M. J. (2015). Working memory, online planning and L2 self-repair behaviour. In Z. Wen, M. Borges & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 160-174). Bristol, England: Multilingual Matters.

Allport, A. (1988). What concept of consciousness? In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in contemporary science* (pp. 159-182). London, England: Clarendon Press.

Alptekin, C., & Erçetin, G. (2009). Assessing the relationship of working memory to L2 reading: Does the nature of comprehension process and reading span task make a difference? *System, 37,* 627-639.

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

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36* (3): 189-208.

Baddeley, A.D. (2015). Working memory in second language learning. In Z. Wen, M. Borges & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17-28). Bristol, England: Multilingual Matters.

Baralt, M. (2015). Working memory capacity, cognitive complexity and L2 recasts in online language teaching. In Z. Wen, M. Borges & A. McNeill (Eds.), *Working memory in* *second language acquisition and processing* (pp. 248-269). Bristol, England: Multilingual Matters.

Briscoe, J., & Rankin, P. M. (2009). Exploration of a 'double-jeopardy' hypothesis within working memory profiles for children with specific language impairment. *International Journal of Language & Communication Disorders*, *44*, 236 - 250.

Cain, K. (2006). Individual differences in children's memory and reading comprehension: An investigation of semantic and inhibitory deficits. *Memory, 14*, 553–569.

Cain, K., & Bignell, S. (2014). Reading and listening comprehension and their relation to inattention and hyperactivity. *British Journal of Educational Psychology, 84*, 108–124.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology, 62,* 73-101.

Cognitivefun (2008-2012). *Cognitive tests: Color Reading Interference (Stroop)*. Available at http://cognitivefun.net/test/2

Council of Europe (2011). *Common European Framework of Reference for Languages*. Strasbourg: Language Policy Division, Council of Europe. Available at http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

De Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19,* 249–276.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–57). Cambridge, England: Cambridge University Press.

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

Ellis, N. C. (2007). The weak-interface consciousness, and form-focussed instruction: Mind the doors. In S. Fotos & H. Nassaji (Eds.), *Form focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 17-33). Oxford: Oxford University Press.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA & bilingualism* (pp. 35-47). Washington, DC: Georgetown University Press.

Ellis, N.C., & Sinclair, S.G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49A*, 234-250.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*, 83-107.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis (Ed.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Bristol, England: Multilingual Matters.

Erçetin, G., & Alptekin, C. (2013). The explicit/implicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied Psycholinguistics, 34*, 727-753.

Erickson, A.P. (2013). The ABC's of font: Effects of changing default fonts. *EWU Masters Thesis Collection.* Paper 108.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*, 147–171.

Gathercole, S., & Baddeley, A. D. (1993). *Working memory and language* (Essays in cognitive psychology). Mahwah, NJ: Lawrence Erlbaum.

Gathercole, S. E., Willis, C. S., Baddeley, A. D. & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory, 2*, 103-127.

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

Godfroid, A.**,** & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 73-90). Philadelphia, PA: John Benjamins.

Grey, S., Williams, J.N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences, 38,* 44–53.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14,* 25-38.

Hassin, R. R., Bargh, J. A., Engell, A. D., & McCulloch, K. C. (2009). Implicit working memory. *Consciousness and Cognition, 18,* 665–678.

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). *Range and frequency programs*. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation.aspx.

Hull R., Martin, R.C, Beier, M.E., Lane D, & Hamilton A.C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology, 22,* 508–522.

Hulstijn, J. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning. The virtues and pitfalls of a two system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25-47). Amsterdam: John Benjamins.

Issa, B. Morgan-Short, K. Villegas, B. & Raney, G. (2015). An eye-tracking study on the role of attention and its relationship with motivation. In L. Roberts, K. McManus, N. Vanek & D. Trenkic (Eds.) *EUROSLA Yearbook 2015* (pp. 114-142). Amsterdam: John

Benjamin Publishing Co.

James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt.

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia*, *10*, 196-214.

Jonides, J., & Smith, E.E. (1997). The architecture of working memory. In M.D. Rugg (Ed.), *Cognitive Neuroscience* (pp. 243-276). Sussex, England: Psychology Press.

Juffs, A. & Harrington, M.W. (2011). Aspects of working memory in L2 Learning. *Language Teaching: Reviews and Studies*, *42*, 137-166.

Koch, C. & Tsuchiya, N. (2006). Attention and consciousness: Two distinct brain processes. *TRENDS in Cognitive Sciences*, *11*, 16-22.

Lamme , V. A. F . (2003). Why visual attention and awareness are different. *TRENDS in Cognitive Sciences*, *7*, 12-18.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*, 202–220.

Li, S. (2015). Working memory, language analytical ability and L2 recasts. In Z. Wen, M. Borges & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 139-159). Bristol, England: Multilingual Matters.

Linck, J.A., Osthus, P., Koeth, J.T., & Bunting, M.F. (2013). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, 1-23.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S.R., Smith, B. K., Bunting, M. F., & Doughty, C. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530-566.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-210). Amsterdam, The Netherlands: John Benjamins.

Martin, K.I., & Ellis, N.C. (2012). The roles of phonological short-term memory and working memory in l2 grammar and vocabulary learning. *Studies in Second Language Acquisition 34*, 379– 413.

Martini, M., Sachse, P., Furtner, M. R., & Gaschler, R. (2015). Why should working memory be related to incidentally learned sequence structures. *Cortex*, *64*, 407-410.

Max Velmans, G. (2009). How to define consciousness—and how not to define consciousness. *Journal of Consciousness Studies 16*(5), 139-156.

McFarland, D. (2006). *Dictionary of animal behavior*. Oxford: Oxford University Press.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex ‘‘frontal lobe’’ tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

Miyake, A. & Friedman, N.P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy, & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.

Miyake, A. & Friedman, N.P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8-14.

Monsell, S. (2005). The chronometrics of task-set control. In J. Duncan, L. Phillips & P. McLeod (Eds.), *Measuring the mind: Speed, control, and age* (pp. 161–90). Oxford: Oxford University Press .

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review 83*, 435-450.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

O'Brien, I., Segalowitz, N., Collentine, J. & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics 27*, 377–402.

Pett, M., Lackey, N., & Sullivan, J. (2003). *Making sense of factor analysis*. Thousand Oaks, NJ: Sage.

Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children*. London, England: Psychological Corporation.

Puric, D., & Pavlovic, M. (2012). Executive function of shifting: Factorial structure and relations to personality and intelligence domains. *Suvremena Psihologija*, *15*(2), 177-191.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219-235.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 888-896.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning, 62*, 93–132.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the saccade task. *Journal of Experimental Psychology: General, 123,* 374–393.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition*, *27,* 235–268.

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.). *The Routledge handbook of second language acquisition* (pp.247-267). London: Routledge.

Rogers, R.D., & Monsell S. (1995). The cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124,* 207–231

Sagarra, N., & Abbuh, R. (2013). Optimizing the noticing of recasts via computer delivered feedback: Evidence that oral input enhancement and working memory help second language learning. *The Modern Language Journal, 97,* 196-216.

Santamaria, K., & Sunderman, G. (2015). Working memory in processing instruction: The acquisition of French clitics. In Z. Wen, M. Borges & A. McNeill (Eds.). *Working memory in second language acquisition and processing* (pp. 205-223). Bristol, UK: Multilingual Matters.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics 11*, 129–158.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker, *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

Shahabi S. R., Abad F. J., & Colom R. (2014). Short-term storage is a stable predictor of fluid intelligence whereas working memory capacity and executive function are not: A comprehensive study with Iranian schoolchildren. *Intelligence, 44*, 134–141.

Shiffrin, R.M. (1988). Attention. In R.A. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Luce (Eds.), *Stevens' handbook of experimental psychology Vol. 2 Learning and Cognition* (pp.739-811). New York, NY: Wiley.

Shin-Cunningham, B.G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Science 12*, 182–186.

Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*, 367-384.

Smith, E.E., & Kosslyn, S.M. (2006). *Cognitive psychology: Mind and brain*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Snow, R. E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist*, *27*, 5-32.

Soto, D., & Silvanto, J. (2014). Reappraising the relationship between working memory and conscious awareness. *Trends in Cognitive Sciences, 18,* 520-525.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263-308.

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30,* 79 – 99.

St Clair-Thompson, H.L., & Gathercole, S.E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*, 745-59.

Styles, E. A. (2006). *The psychology of attention 2nd edition*. Hove, NJ: Psychology Press.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn and Bacon.

Tagarelli, K.M., Borges Mota, M., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. In L. Carlson, C. Holscher & T. Shipley (Eds.). *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2016-2066). Austin, TX: Cognitive Science Society.

Tagarelli, K.M., Borges Mota, M., & Rebuschat, P. (2015). Working memory, learning conditions and the acquisition of L2 syntax. In Z. Wen, M. Borges & A. McNeill (Eds.). *Working memory in second language acquisition and processing* (pp. 224-247). Bristol, England: Multilingual Matters.

Tamnes, C.K., Walhovd, K.B., Grydeland, H., Holland, D., Østby, Y., Dale, A.M., & Fjell, A.M. (2013). Longitudinal working memory development is related to structural maturation of frontal and parietal cortices. *Journal of Cognitive Neuroscience, 25*, 1611–1623.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33,* 285-318.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between operation span and raven. *Intelligence, 33,* 67– 81.

Van der Linden, M., Brédart, S., & Beerten, A. (1994). Age-related differences in updating working memory. *British Journal of Psychology, 85*, 145–152.

Ward, G., & Allport, A. (1997). Planning and problem-solving using the 5-disc Tower of London task. *Quarterly Journal of Experimental Psychology*, *50*, 49-78.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Personality Science and Individual Differences, 4,* 433.

Williams, J.N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53,* 67-121

Wolfe, J.M., & Horowitz, T.S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews 5,* 1-7.

Yilmaz, Y. & Granena, G. (2015). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*. doi:10.1017/S136672891400090X

Zheng, D., Dong, X., Sun, H., Xu, Y., Ma, Y., & Wang, X. (2012). The overall impairment of core executive function components in patients with amnestic mild cognitive impairment: A cross-sectional study. *BMC Neurology*, *12*, 138.

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *31*, 3–29.

Table 1

Number of participants whose eye-tracking data was of sufficiently good quality in each stage and the overall number of participants included in the eye-tracking data analyses

	Text 1	Text 2	Text 3	No. of participants included in the analysis
enhanced+ instructions	18	16	13	10
enhanced+ instructions+ explanation	17	16	15	14
enhanced only	15	16	17	11
unenhanced	12	14	14	10
percentage	77.5%	77.5%	73.75%	56.25%

Table 2

Cognitive variables: Descriptive statistics and comparisons of treatment groups

Group	Digit span Mean (SD)	Keep track Mean (SD)	Stroop Mean (SD)	Plus Minus Mean (SD)	Composite WM factor score Mean (SD)
enhanced+ instructions	5.40 (1.35)	77.40 (13.53)	2127.30 (605.45)	18.75 (17.11)	.064 (.922)
enhanced+instructions+ explanation	5.70 (1.72)	75.73 (16.92)	1822.15 (448.99)	23.90 (19.62)	.258 (1.02)
enhanced only	5.20 (1.50)	73.51 (14.28)	2131.50 (533.89)	15.70 (10.13)	-.093 (.999)
unenhanced	5.05 (1.66)	70.21 (17.26	2150.15 (892.80)	17.75 (13.47)	-.229 (1.958)
F	.641	.796	1.201	1.012	.880
p	.591	.500	.315	.392	.455

Table 3

Correlations among the cognitive variables

	Digit span	Keep-track	Plus Minus	Stroop
Digit span		.818**	.112	-.530**
Keep-track			.119	-.455**
Plus Minus				-.069

**. Correlation is significant at the 0.01 level

Table 4

Descriptive statistics for the SR and GJ tasks

Task	Groups (n= 20)	Mean Pretest (SD)	Mean post-test (SD)	Mean gain (SD)
SR task	*control*	0.00 (0.00)	0.35 (0.93)	0.35 (0.93)
	enhanced+ instructions	0.05 (0.22)	1.60 (1.81)	1.55 (1.70)
	enhanced+ instructions+ explanation	0.05 (0.22)	2.20 (1.54)	2.15 (1.42)
	enhanced only	0.00 (0.00)	0.60 (0.99)	0.60 (0.99)
	unenhanced	0.05 (0.22)	0.35 (0.48)	0.30 (0.47)
	Total for experimental groups	0.03 (0.07)	1.19 (1.49)	1.15 (1.42)
GJ task	*control*	4.75 (0.42)	4.85 (0.45)	0.10 (1.44)
	enhanced+ instructions	4.25 (1.61)	6.05 (2.28)	1.80 (1.85)
	enhanced+ instructions+ explanation	4.85 (1.42)	6.85 (2.08)	2.00 (1.07)
	enhanced only	4.25 (1.29)	5.75 (1.88)	1.50 (1.57)
	unenhanced	4.65 (1.18)	5.50 (1.39)	.85 (1.34)
	Total for experimental groups	4.50 (1.38)	6.04 (1.97)	1.54 (1.32)

Table 5

Spearman rank order correlations between gains in the SR and GJ task and the composite WM and the Plus Minus task scores

Group	n	SR gain score Plus Minus	GJ gain score task	SR gain Composite	GJ gain WM
whole experimental sample	80	.070	-.024	.614**	.658**
enhanced+ instructions	20	-.490*	-.307	.701**	.524*
enhanced+ instructions+ explanations	20	.320	.159	.887**	.770**
enhanced only	20	.264	-.176	.533*	.714**
Unenhanced	20	.103	.292	.454*	.619**

*Correlation significant at $p < 0.05$ level

**Correlation significant at $p<0.01$ level

Table 6

Descriptive statistics for the eye-tracking measures

	Group (n= 20)	Mean	SD
Composite TFD	*enhanced+ instructions*	2.35	0.53
	enhanced+ instructions+ explanation	2.80	0.74
	enhanced only	1.50	0.21
	unenhanced	.99	0.32
	total	1.98	0.88
Total DTFD	*enhanced+ instructions*	4.45	1.78
	enhanced+ instructions+ explanation	5.48	2.15
	enhanced only	1.51	0.91
	unenhanced	0.00	0.30
	total	3.06	2.69

Table 7

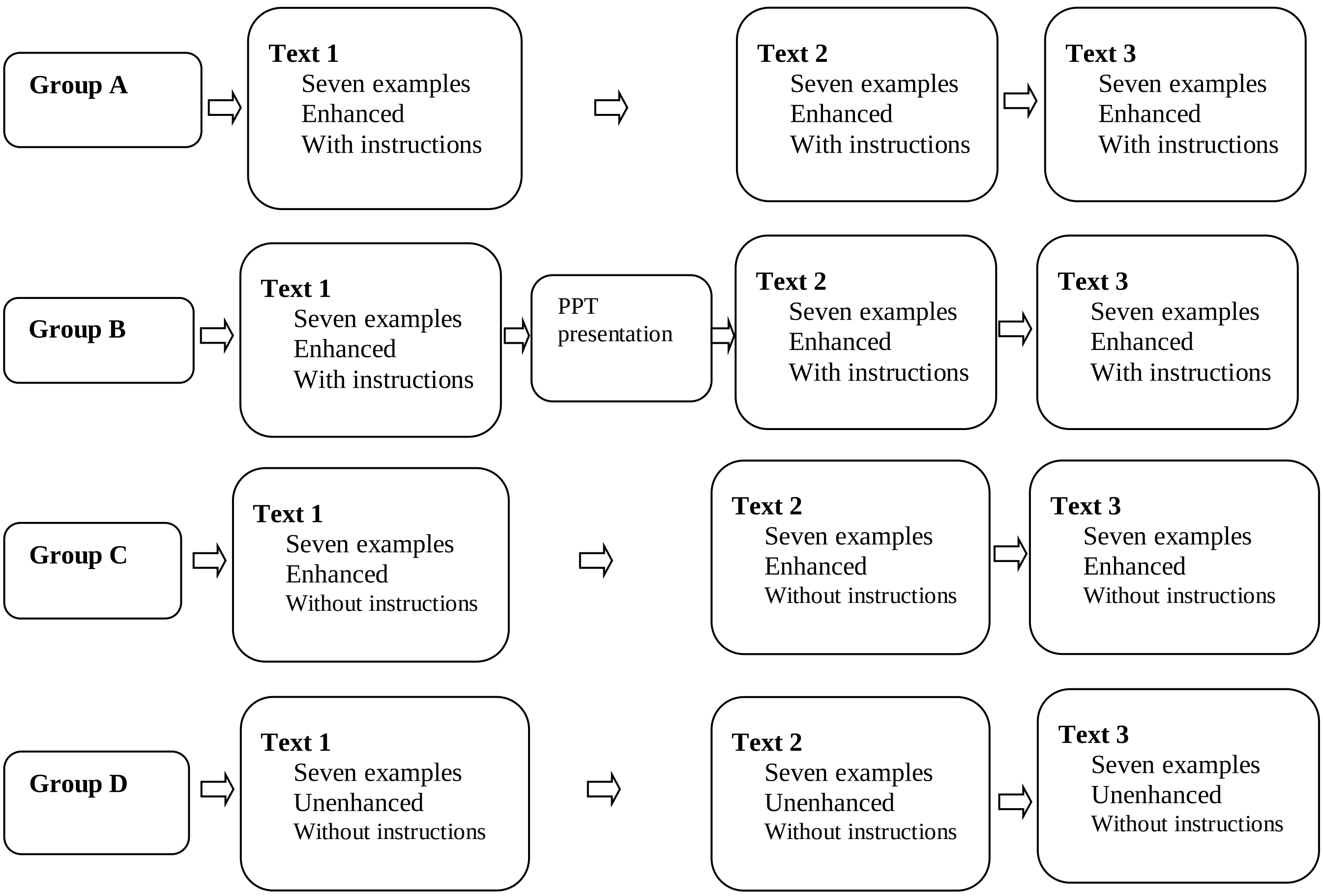
Correlations between eye-tracking measures and GJ task and the composite WM and Plus Minus task score

Group	*n*	TFD Plus	ΔDOE Minus	TFD Composite	ΔDOE WM
whole experimental sample	45	.052	.030	.324*	.394**
enhanced+ instructions	10	-.455	-.333	.382	.697*
enhanced+ instructions+ explanations	14	.055	-.007	.859**	.824**
enhanced only	11	.064	-.334	.155	.673*
unenhanced	10	-.405	-.203	-.455	-.030

*Correlation significant at 0.05 level

**Correlation significant at p<0.01 level

Figure 1. The design of the study



Group A: enhanced + instructions
Group B: enhanced+ instructions+ explanation
Group C: enhanced
Group D: unenhanced

Figure 2. Example of the unenhanced input slides

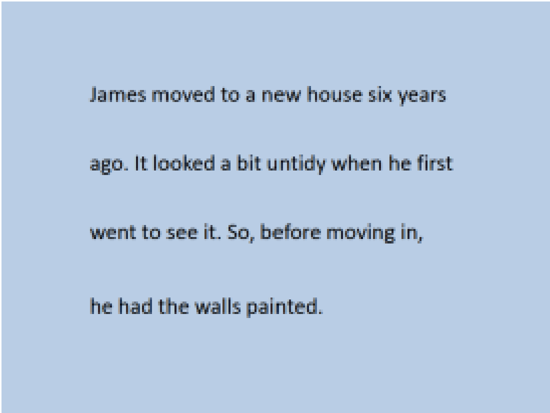


Figure 3. Example of the enhanced input slides

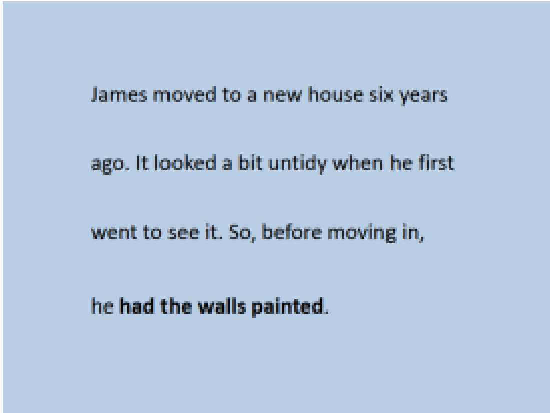
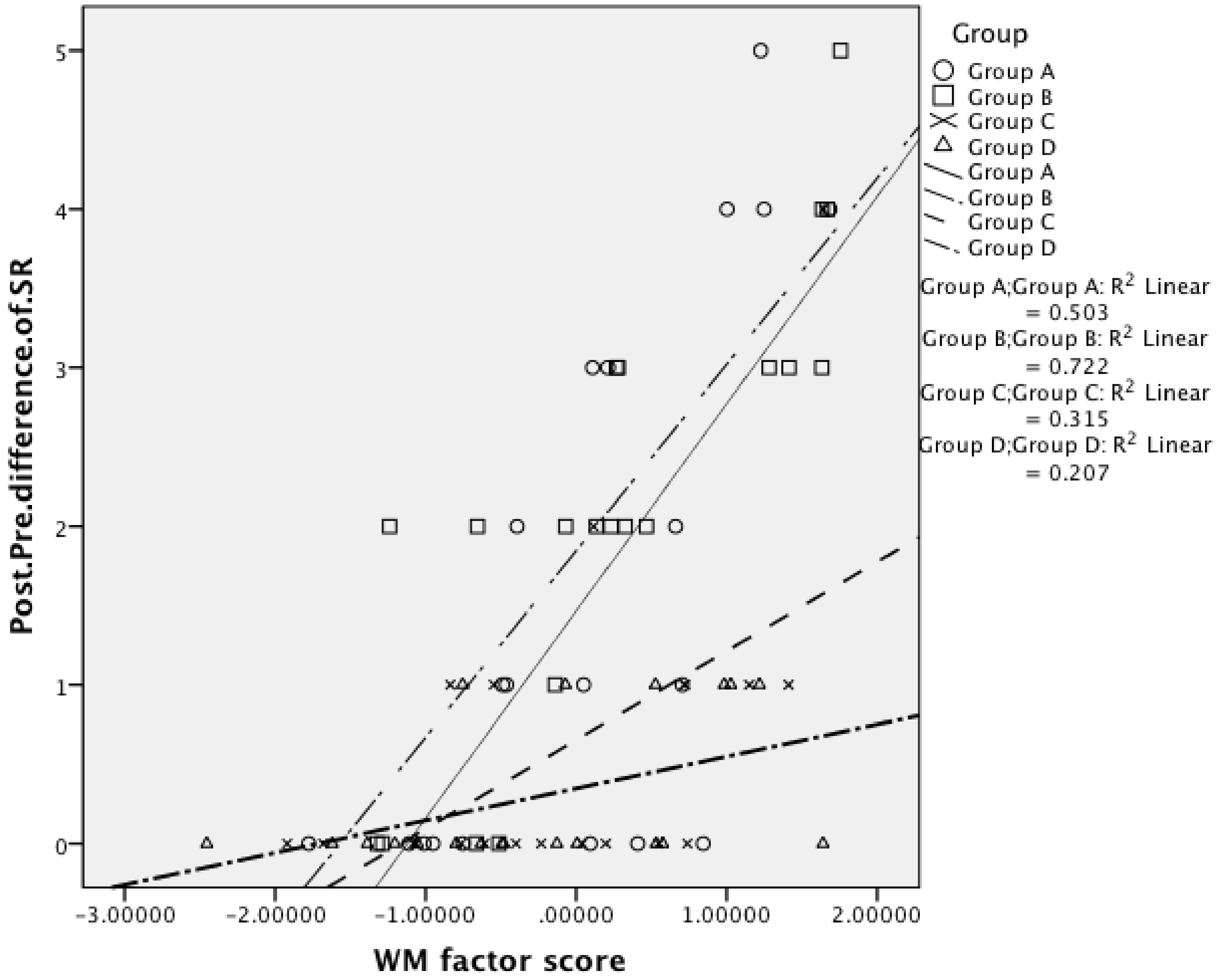
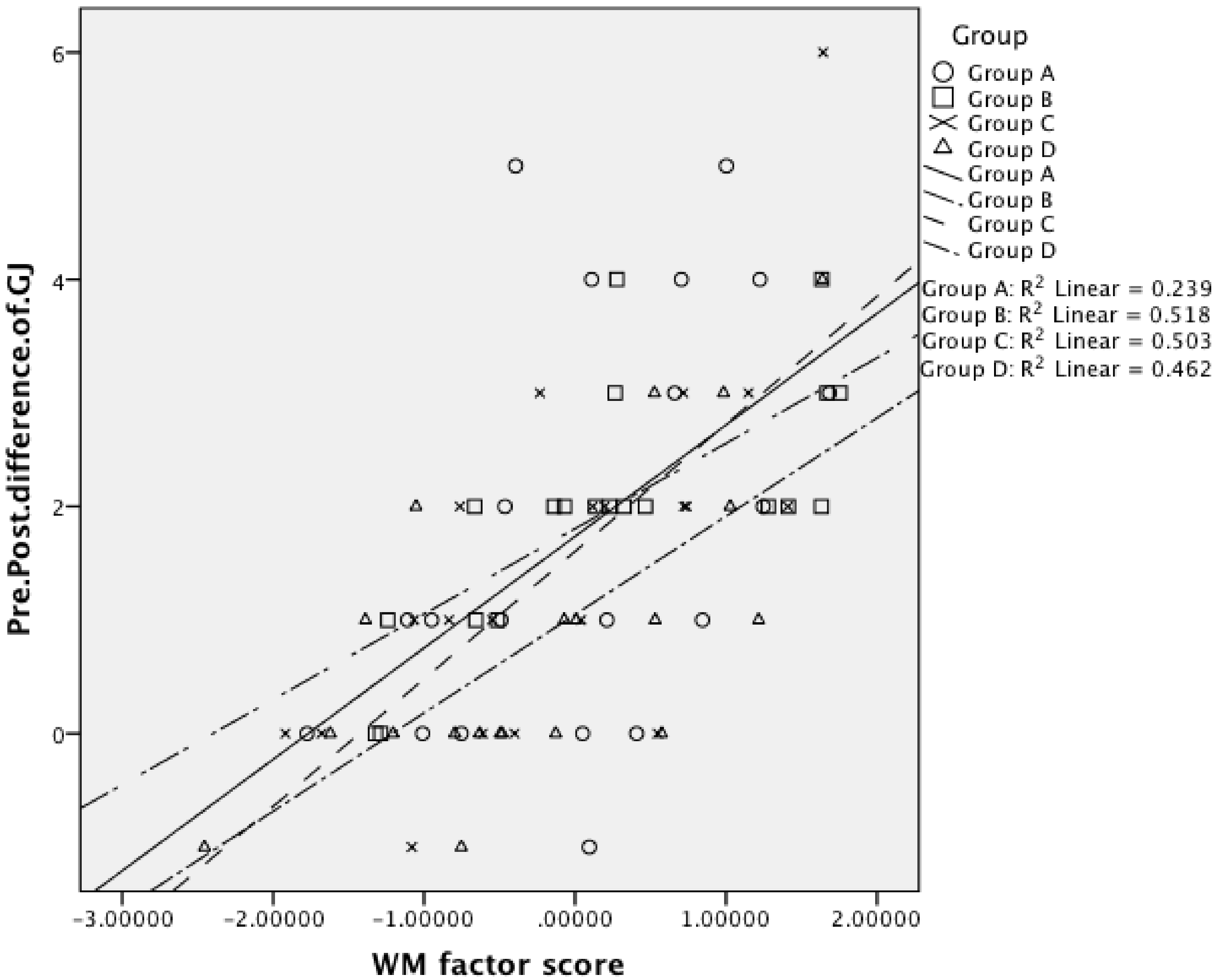


Figure 4. Interaction between WM capacity and instructional condition in the SR task



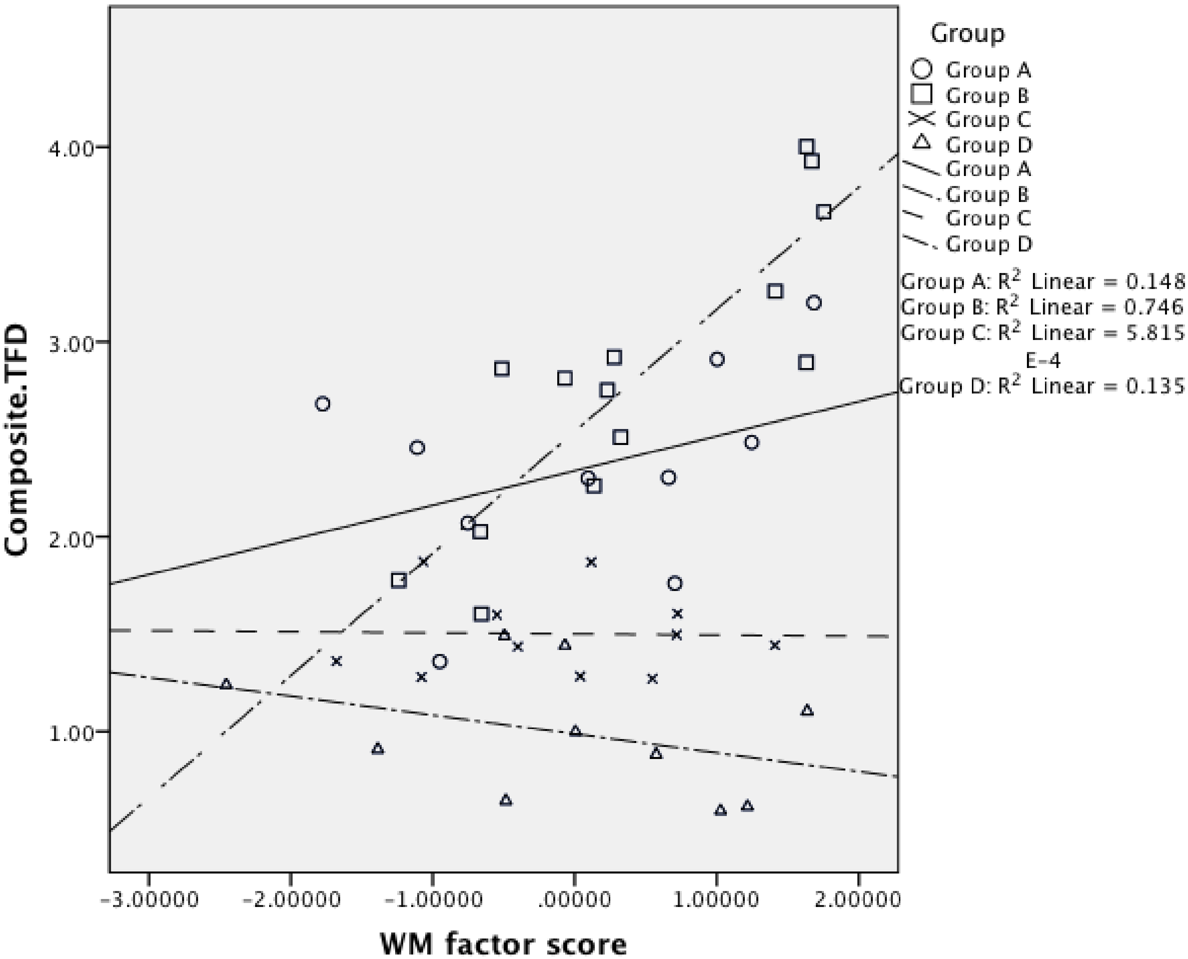
Group A: enhanced + instructions
Group B: enhanced+ instructions+ explanation
Group C: enhanced
Group D: unenhanced

Figure 5. Interaction between WM capacity and instructional condition in the GJ task



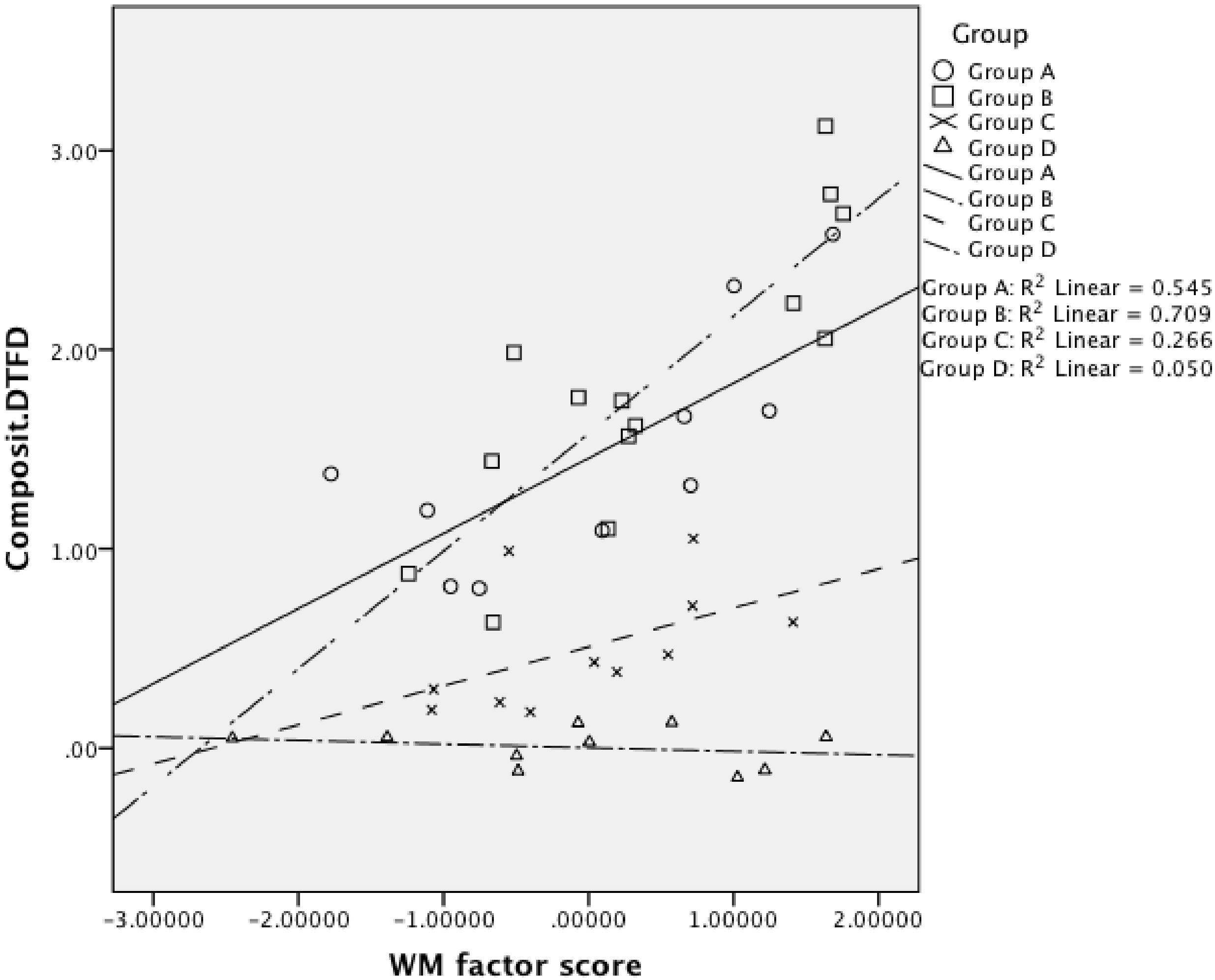
Group A: enhanced + instructions
Group B: enhanced+ instructions+ explanation
Group C: enhanced
Group D: unenhanced

Figure 6. Interaction between WM capacity and instructional condition in the case of the TFD measure



Group A: enhanced + instructions
Group B: enhanced+ instructions+ explanation
Group C: enhanced
Group D: unenhanced

Figure 7. Interaction between WM capacity and instructional condition in the case of the DOE measure



Group A: enhanced + instructions
Group B: enhanced+ instructions+ explanation
Group C: enhanced
Group D: unenhanced