

An Experimental Study Towards Investigating the Effect of Working Memory Capacity on Complex Diagram Understandability

Nergiz Sözen¹, Bilge Say², Özkan Kılıç³

¹*Atilim University Department of Computer Engineering, Golbasi, Ankara, Turkey*

²*Atilim University Department of Software Engineering, Golbasi, Ankara, Turkey*

³*Yildirim Beyazıt University Department of Computer Engineering, Yenimahalle, Ankara, Turkey*

Abstract –This study investigates whether working memory (WM) capacity affects the understandability of complex diagrams and if so, whether WM training has a positive effect on their comprehensibility. Two experiments were conducted with computer science students. In the first experiment, we collected eye-tracking data while participants performed comprehension tasks on an activity diagram. In the second experiment, the participants completed WM training, before and after their comprehension scores were measured. The results showed that working memory capacity can positively affect the understandability of complex diagrams, but it provided no conclusive evidence for the effectiveness of working memory training.

Keywords – cognitive training, visual search, working memory capacity, diagram comprehension.

1. Introduction

Understanding complex diagrams is an important concern because as diagram complexity increases,

diagram understandability decreases especially for novice modelers [1]. To perform comprehension tasks on a diagram efficiently and correctly, the diagram should be understood by all stakeholders such as developers, analysts, business process modelers in an information systems development setting. Various factors affect the understandability of diagrams, commonly specified under the categories of model-related issues and personal issues [2]. This study focuses on one of the personal issues: working memory capacity. The positive effect of working memory improvement on various cognitive tasks such as problem solving, and fluid intelligence is controversial in the literature. A number of research studies present empirical findings that skills gained with proper working memory training can be transferred to other cognitive tasks that are different than the training. There are also studies suggesting no evidence of any positive effect of working memory training on improved performance for other cognitive tasks [3], [4], [5]. Jaeggi and her colleagues found no evidence that a working memory training can improve fluid intelligence [6]. Although there are a number of research studies in the literature that investigate the effect of personal factors in model understandability, to the best of our knowledge there are no studies that investigate whether working memory capacity and training have a positive effect on diagram understandability.

This paper investigates the effect of working memory capacity on activity diagram and ER diagram comprehension by providing empirical evidence gathered from novice computer science students with experiments using cognitive and performance measures as well as eye-tracking data. This paper tries to answer the following research questions:

Research Question 1: How does working memory capacity affect the comprehension skills and eye

DOI: 10.18421/TEM94-09

<https://doi.org/10.18421/TEM94-09>

Corresponding author: Nergiz Sözen,
Department of Computer Engineering Atilim University,
Ankara, Turkey.


Email: kilinc.nergiz@gmail.com

Received: 02 October 2020.

Revised: 01 November 2020.

Accepted: 06 November 2020.

Published: 27 November 2020.

 © 2020 Nergiz Sözen, Bilge Say & Özkan Kılıç;
published by UIKTEN. This work is licensed under the
Creative Commons Attribution-NonCommercial-NoDerivs
4.0 License.

The article is published with Open Access at
www.temjournal.com

movements of the novice modelers given a UML activity diagram comprehension task?

Research Question 2: Does working memory capacity training have a positive effect on the entity-relationship diagram understandability and working memory capacity?

We conducted two experiments to investigate the effect of working memory on diagram comprehension 1) by collecting eye-tracking data of the participants and measuring their performance on a comprehension task about an activity diagram and 2) by providing working memory training to the participants and measuring their performance on the entity-relationship diagrams before and after the training.

The first experiment was correlational and exploratory in nature while the second one involved an experimental manipulation.

2. Literature

Business process models and data models provide a detailed abstract visual representation of the organizational procedures to ease the understandability of the complex work that is needed to be performed by any organization [7]. Business processes are also considered a substantial artifice that promises to sustain and increase enterprise performance and rivalry [8]. Usually, complex systems are composed of smaller tasks and activities that have to be performed to achieve the goals of organizations. Deconstructing complex systems into a web of smaller processes is crucial for successful implementation of complex information systems. There are multifold ways to visually represent the interrelated business processes. In this study we preferred activity diagrams to represent business processes. Activity diagrams are similar to flow diagrams which show the flow from activity to activity and can be used to represent complex business processes. Additionally, we used entity-relationship diagrams (ERD) as data models to investigate how working memory training affects comprehension of different diagrams used in software industry. ER diagrams are visual representations of databases that are used to illustrate the major entities in a system and the inter-relationships among these entities. ERD models are also used to represent a large complex system in smaller entities that are expected to be more manageable and easier to comprehend. Although one of the main purposes of all these visual representations of complex businesses is to ease the understandability of the interactions between processes, the models may become difficult to understand, especially for novice modelers. There are empirical indications that the understandability of

business process models hinges upon a diversity of factors, commonly specified under the categories of model-related issues and personal issues [2] also referred as human factors [9]. There are also personal factors under the human factor category that affect Business Process Model understandability. The most-researched human factors are professional background, cognitive abilities, learning style, motive and strategy, and domain familiarity. In this experimental study we investigate the effect of one of the personal factors: working memory capacity on the Business Process Model and Data Model understandability. The literature presents a number of empirical studies that investigate the effect of cognitive factors -- in particular cognitive load theory -- on model understandability where the studies do not consider working memory capacity. Working memory capacity is a factor that is directly related to the cognitive load, yet, to the best of our knowledge, there are no studies that specifically focus on working memory capacity. In this paper, one of the cognitive factors i.e., working memory capacity, is investigated because of the insufficient amount of research that has investigated its effect on Business Process Model and data model understandability.

Working memory is a limited capacity system of the brain that operates, organizes, contrasts, compares and processes information on complex cognitive tasks [10]. Limitations in working memory capacity make it challenging to process information and perform cognitively demanding activities such as problem solving, reading, and learning. Since this finite capacity is a genetic human characteristic, it can vary among individuals and affect cognitive abilities such as learning or information processing. Working memory capacity plays a great role in learning tasks especially the ones that require high cognitive effort such as reasoning, verbal comprehension, multitasking, verbal fluency, maths and problem solving [11]. For this reason, performing such tasks can often be challenging, which depend on the capacity of working memory. Despite its genetic nature, working memory capacity can be improved with proper cognitive trainings, according to the literature [6],[12],[13]. Some of these cognitive trainings include corsi block tapping [14],[15], and span tasks such as counting span, operation span and reading span [16]. Among the various tasks described in the literature that promise to improve working memory capacity, we selected three of the most commonly used: AOSpan [17], Symmetry Span [17], and dual n-back [6]. All these working memory tasks require high cognitive performance. We used AOSpan and Symmetry Span tasks together to measure complementary aspects of our participants' working memory capacity. AOSpan is a complex verbal task where participants need to

remember a sequence of letters and perform arithmetic operations simultaneously. Symmetry Span is more of a visual task in which participants need to remember the position of red cells on a grid while they are expected to mark whether a given figure is symmetrical or asymmetrical. We also used adaptive dual n-back task to train the participants in our experimental group during the second experiment. Dual N-back training expects participants to remember n-back position and audio and decide whether they match with the current position and audio (See Figure 2).

There are studies investigating Business Process Model understandability using eye-tracking method [18], [2], [19]. In our study, we also collected eye-tracking data from our participants for triangulation purposes, to further elaborate, understand and interpret our findings gathered from working memory and comprehension tests in the first experiment. Eye-tracking data of novice modelers with high vs. low working memory capacity as well as the modelers who answered the comprehension tasks correctly vs. incorrectly were compared.

3. Research Method

As mentioned earlier, we have conducted two different experiments to understand the effect of working memory capacity on the Business Process Model (BPM) and data model understandability. As for the representation of Business Process, we used a UML activity diagrams. For the representation of data model, we used an entity-relationship (ER) diagram.

In the first experiment, we measured working memory capacity of the participants, and then had them solve a computerized comprehension test about a specific BPM represented in the form of activity diagrams. This activity diagram was the representation of the process of complaint handling in a large organization (See Figure 1). While the participants were performing the computerized test, we recorded their raw eye-tracking data using Tobii Eye-Tracker 4C 90Hz. In the second experiment, we have designed a quasi-experiment to investigate whether the participants' ER diagram understandability would increase after performing seven weeks of working memory training. Details of both experiments are presented separately under the Experiment Procedures section.

3.1. Participants

The first experiment was conducted with 48 third-year undergraduate students who were enrolled in SE 346 Software Engineering course, in which they learn UML Activity Diagrams for the first time.

Therefore, familiarity with Activity Diagrams (ACDs) as well as educational status were similar for all the participants. To ensure that all the participants have approximately the same levels of familiarity, we tested the participants before the experiment with a familiarity test about UML activity diagram notation. Although 48 students participated in the first experiment, one of the participants who scored extremely low on the activity diagram familiarity test was been eliminated. After the elimination, the mean familiarity score for 47 participants was ($M = 57.31$, $SD = 14.47$).

We conducted our second experiment with 72 third-year undergraduate students who were taking CMPE 341 Database Design and Management course where they learn ER diagrams for the first time. However, 32 of these participants abandoned the experiment at different stages. After the post-training assessment, the data was pre-processed, and 24 participants were eliminated as their working memory capacity accuracy scores were lower than 75% threshold [16]. This threshold is used to ensure that the participant is not using the time allocated to perform distraction operations such as simple arithmetic operations or symmetry of the figures for the memorization task.

We incentivized the experiment participations with three bonus points over 100 for each experiment.

3.2. Materials

This section presents the materials we used in our experiments.

In our first experiment we used the following materials:

- Activity Diagram (ACD) Familiarity Test: a true-false type 16-question test to measure the participants' familiarity with activity diagrams.
- Working Memory Span Tasks: AOSpan and Symmetry Span tasks [17] for measuring working memory capacity.
- Computerized (ACD) Understandability Test: an 8-multiple-choice computerized test to measure understandability of activity diagrams. The activity diagram was adapted from the BPMN diagrams used in [20] with the authors' permission (See Figure 1). The diagram represents the process of complaint handling in an organization. The diagram that the participants see remains on the left while different questions appear on the right (See Figure 1).
- Eye-tracker device: Tobii Eye-Tracker 4C, 90 Hz.

We used the following materials for the second experiment:

- Entity-Relationship Diagram (ERD) Familiarity Test: A true-false type 13-question test to measure ERD familiarity of the participants.
- Working Memory Span Tasks: AOSpan, Symmetry Span tasks to measure working memory capacity of the participants.
- Entity-Relationship Diagram (ERD) Understandability Tests: Two paper-pencil tests with 19 true-false questions about a Small Airport Business Process Model and Railway Systems to measure Business Process Model Understandability using ERD language. The small airport ERD was prepared to measure data model understandability in pre-training (before working memory training) and railway systems to measure BPM understandability in post-training (after working memory training) [21].
- Adaptive Visual Search: We have developed an Android application for the active-control group to perform a dummy treatment that has no reported effect on either improving working memory capacity or other cognitive tasks [3]. Using an active-control group was important for this research because it helped us consider the possibility that improvement within the experimental group resulted from a placebo effect. Visual search tasks expect the participant to detect whether a target figure is presented among a number of distractor figures. Adaptive visual search application allows the experimenter to select the number of distractor figures and the images that experimenter wants to use for the target and distractors. The experimenter can also determine the number of sessions and the minimum score in order to successfully complete a block.
- Adaptive Dual n-back: Participants in the experimental group used the mobile version of dual n-back training which was equivalent to the dual n-back training that Jaeggi and his colleagues [6] offered. Each session comprised ten blocks of 20 trials that the participants had to complete. On each trial, the participants were presented a visual target location and an audio target letter simultaneously. Figure 2 is an example of dual 2-back task where the participant needs to 1) memorize and compare the current audio target letter and position of visual target with the 2-back audio target and visual target position on each trial and 2) select the match-type as either letter match, audio match, or both. As seen in Figure 2, On the 3rd trial auditory target letter is "C" and the position of visual target is the right bottom of the grid. Because the n-level is 2 for that example, participant have to check the 2-back visual and audio target. Visual and auditory targets are both the same with the ones in 2-back trial. Therefore, the participant has to state that both targets match. The adaptive dual n-back training starts with dual 1-back. After completing each block, participants receive feedback on their dual n-back performance. If the participant's response is at least 90% accurate for both audial and visual parts, the n-level adapts and increases in the next block. If a participant's response is 70% or less accurate, the n-level decreases in the next block. Otherwise, the participant's n-level remains the same in the next block. We have used a mobile application for the entire dual n-back training which was developed by "IQ Mindware" [22].

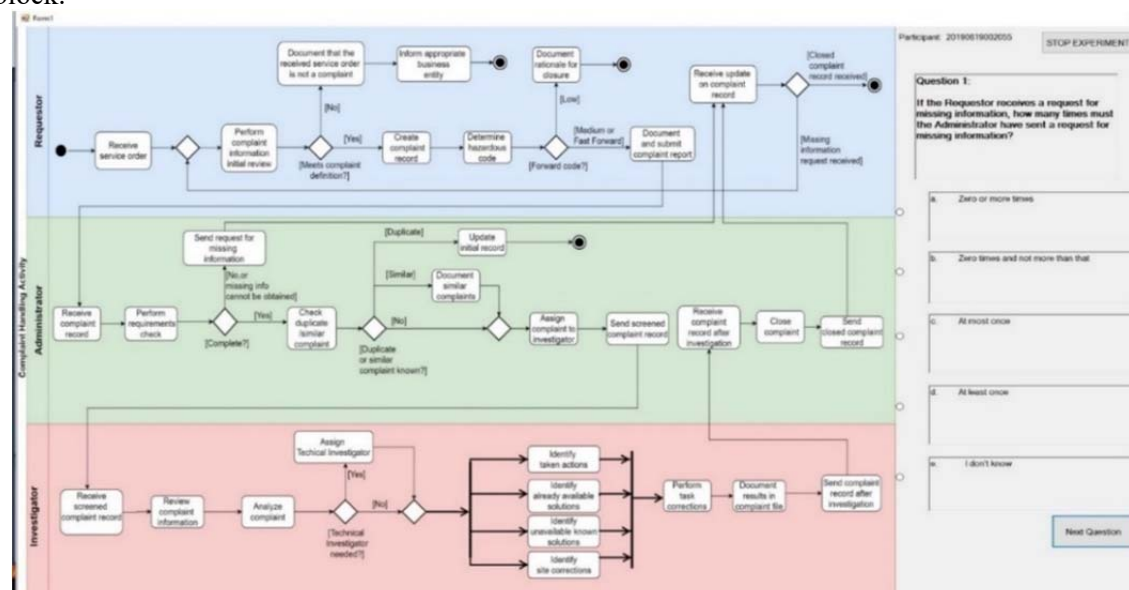


Figure 1. A sample question from Computerized ACD Understandability Test

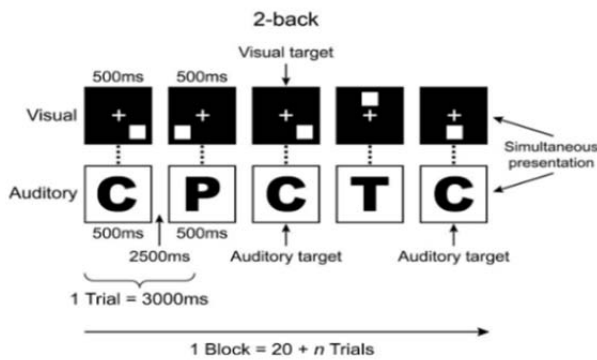


Figure 2. Dual n-back Task [6]

3.3. Experiment Procedures

This section presents the detailed procedures for both experiments. We followed the procedure approved by the Ethical Committee of Atılım University for both experiments.

The purpose of conducting the first experiment was to understand if working memory capacity has an effect on business process model understandability for the inexperienced novice modelers. Each participant participated in the experiment individually and were informed that the experiment was voluntary and could be left any time on demand. All the necessary instructions and forms were provided to the participants when the participant arrived. First, each participant solved the ACD familiarity test, after which the computerized ACD understandability test started. Participants were provided the necessary instructions before starting the test. Then, the eye tracker calibration was performed to make sure that the head and eye positions were aligned with the display and eye tracker hardware. After successful calibration, a two-question warm-up test was presented to prevent any distorting effect of eye-tracking. Then, the computerized eight-question ACD understandability test was performed by the participants. The participant was then given a two-minute break, after which the participant completed the AOSpan and Symmetry Span tasks.

In the first experiment, various eye metrics listed below were calculated using Python pyGaze library in order to measure the eye movement. The following definitions are important for understanding the metrics:

- **Fixation Duration:** The total time that the participant's gaze is directed on a specific model element.
- **Number of Fixations:** The total number of times a gaze fixated on a model element.
- **Area of Interest (AOI):** A specific region on model that might include one or more elements of a model that the participants need to focus on in order to solve a problem.

The followings are the calculated eye-tracking metrics and other metrics for the first experiment:

- Total number of fixations [23].
- Total duration of fixations [23].
- Average duration of fixations [23].
- Total number of fixations on AOIs [23].
- Total duration of fixations on AOIs [23].
- Average duration of fixations on AOIs [23].
- **Scan Path Precision (SPP):** This metric provides information about the percentage of fixations on area of interest in relation to all fixations of the participant. This metric tells us to what degree the participant was distracted by elements unrelated to the current question [24].
- **Scan Path Recall (SPR):** This metric provides information about the ratio of fixations on areas of interest to the number of areas of interest for the current [3].
- **Response Time:** This metric describes the time interval between viewing the question and answering the question [23].
- **Return time (on AOI) [23].**
- **Correctness:** This measure indicates whether the participant's answer to the ACD question was correct or incorrect [25]. If the question was answered correctly, the metric is marked as 1; otherwise, 0.
- **Working Memory Performance:** AOSpan score of the participants [26]. According to our analysis, the symmetry span score was not statistically significant on any other metrics or tasks. Therefore, for the rest of the paper, we consider only AOSpan score as working memory capacity. A nominal variable is generated by splitting the AOSpan scores in half and marking the higher WM performances as 1, and lower WM performances as 2.

The second experiment was also based on voluntary participation. Each participant was informed about the instructions of the experiment beforehand. Our aim was to investigate whether extensive, proper working memory capacity (WMC) training has any effect on the Business Process Model and data model understandability, and on the working memory capacity. The second experiment was composed of three phases. For this experiment, we measured the working memory capacity and ERD understandability scores at two point in time: pre-training (in the first phase of the experiment before the treatment), and post-training (in the third phase of the experiment, after the treatment). The detailed experiment procedure for each of the phases is provided in the following section.

First phase: We gathered demographic data from the participants. We then tested all the participants' familiarity levels using an ERD familiarity test in a

laboratory setting. Finally, after the familiarity test, all the participants took the AOSpan and Symmetry Span tasks in the laboratory setting.

Second phase: Based on the ERD familiarity scores of the participants and their devices' operating systems, we formed the control and experimental groups before the second phase. We first assigned the participants to groups based on their mobile device operating system, making sure that the familiarity scores were not statistically significantly different between groups. Participants in active control group ($n=15$) and in experimental group ($n=15$) were equal in number and in level of familiarity. An independent-sample t-test was conducted in order to compare the mean differences of ERD familiarity scores between control and experimental groups. The 15 participants who received dual n-back working memory training ($M=76.9$, $SD=14.9$) compared to the 15 participants in the active control group who performed visual search training ($M=69.2$, $SD=15.1$) demonstrated no significant difference on ERD familiarity scores, $t(28)=1.4$, $p=.17$. Participants in the active control group installed the visual search application; participants in the experimental group installed the dual n-back application. In the second phase of the experiment, participants trained their working memory using a smart phone application. This training consisted of 3 sessions a week for 7 weeks. When the training was completed, the third phase started.

Third phase: The third and last phase of the experiment repeated the procedures of phase one in order to measure the working memory capacity of the participants as well as their ERD understandability.

4. Analysis Results

After the experiments were conducted, the data was analyzed using various statistical analysis methods. Since each experiment provided different type of data, the analysis details are provided in detail:

Experiment 1

We collected eye-tracking data of the participants using eye tracker hardware while participants were answering questions on a computerized ACD understandability test. In order to answer each question, participants were expected to fixate on an expert-defined area of interest on an activity diagram [23]. Each question has a different number of AOI that is relevant to solve the question. Question 1 has six AOIs, Question 6 has two AOIs, Question 3 has three AOIs, Question 4 has three AOIs, Question 5 has one AOI, Question 6 has five AOIs, Question 7 has three AOIs, and Question 8 has one AOI. We also calculated the First Fixations on an AOI. Since not every participant's first fixation falls into the same AOI, the number of measurements for each participant is different in our experiment.

We have also calculated the ACD understandability scores and working memory capacity scores (AOSpan and Symmetry Span). Additionally, the raw eye-tracking data including participant id, x and y coordinates of the eye gaze, left and right pupil diameters, timestamp, question view time, question answer time, and selected answer for each question was used to calculate some commonly used eye metrics. These commonly used eye-metrics include total number of fixations, average number of fixation duration, return time on AOI, scan path precision, scan path recall, response time, total and average duration on AOI, which were explained under the experiment procedures in detail. We calculated eye metrics by adapting the functions in the Python pyGaze library. The calculated metrics were written on a .csv document generated by the code and then imported into SPSS together with the working memory test scores and Familiarity with ACD test scores. Due to tracking errors, a total of 352 healthy measurements were comprised of the 44 participants with 8 questions each. Table 1 illustrates the range, standard deviation, mean, and distribution of each metric.

Table 1. Descriptive statistics for metrics

	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>SD</i>	<i>Shapiro-Wilk Normality</i>
Familiarity with ACD	352	23.08	100.00	70.98	17.47	No, $p<.05$
ACD comprehension	352	12.50	100.00	51.46	21.24	No, $p<.05$
Total Fixation Duration	352	1.40	158.60	32.28	24.67	No, $p<.05$
Response Time	352	4.70	179.67	42.15	30.24	No, $p<.05$
Average Fixation Duration	352	.13	.31	.22	.03	No, $p<.05$
Correctness	352	0	1	.51	.50	No, $p<.05$
Scan Path Precision	352	.00	44.00	11.22	9.59	No, $p<.05$
Scan Path Recall	352	.00	100.00	37.16	32.95	No, $p<.05$

According to the descriptive statistics, the percentage of correctly answered questions is 51%, which shows that the participants were novice modelers. Additionally, scan path precision shows that only 11% of the fixations were on an area of interest, while participants fixated on 37% of all areas of interest.

Analysis for Independent Variable Correctness:

To understand if there is a significant mean difference in terms of fixations, scan path recall, and response time between correctly and incorrectly answered questions, we conducted a Kruskal-Wallis H test. Table 2 shows the result of the test.

Table 2. Kruskal-Wallis H Test Results for (IV) Correctness

	χ^2	<i>p</i>	Mean Rank	
			Correct	Incorrect
Total Number of Fixations	7.27	.01	161.68	190.90
Average Duration of Fixations on AOIs	59.59	.00	216.85	133.49
Scan Path Precision	31.91	.00	205.89	144.89
Scan Path Recall	9.01	.00	191.59	159.38
Response Time	9.80	.00	159.38	193.30

To compare correctly and incorrectly answered questions in terms of Return Time on AOI, we conducted Mann-Whitney U test. The results of Mann-Whitney U test indicated that Return Time on AOI in seconds was higher for correct answers (*Mean Rank*= 342.86) than for incorrect answers (*Mean Rank*= 310.68) with *U*= 47736.00, *p*= .03.

Analysis for Independent Variable Working Memory Performance:

To understand if there is a statistically significant mean difference in terms of fixations, scan path precision, scan path recall, and response time between high vs. low working memory performances, we conducted a Kruskal-Wallis H test. Table 3 illustrates the test results.

Table 3. Kruskal-Wallis H Test results for (IV) WM

	χ^2	<i>p</i>	Mean Rank	
			High WM	Low WM
Total Number of Fixations	4.79	.02	188.96	165.15
Average Duration of Fixations on AOIs	9.85	.00	194.51	160.49
Scan Path Precision	10.27	.00	194.91	160.16
Scan Path Recall	8.80	.00	193.17	161.62
Response Time	4.71	.03	188.84	165.25

To compare high vs. low working memory performances in terms of return time on AOI and activity diagram comprehension, we conducted Mann-Whitney U test. The results of Mann-Whitney U test indicated that Return Time on AOI in seconds was higher for high WM performance (*Mean Rank*= 349.56) than for low WM performance (*Mean Rank*= 310.40), *U*= 47404.00, *p*= .01. Similarly, Mann-Whitney U tests indicated that ACD comprehension score was higher for high WM performance (*Mean Rank*= 196.30) than for low WM performance (*Mean Rank*= 158.99), *U*= 12032.00, *p*= .00

Experiment 2

To find an empirical answer for the second research question, we conducted Experiment 2, and tested the following hypotheses:

*H*₀: There is no statistically significant difference between experimental and control groups in terms of diagram understandability scores.

*H*_a: There is a statistically significant difference between experimental and control groups in terms of diagram understandability scores.

In order to investigate the effect of working memory capacity training on the understandability of ER diagrams, we have measured the working memory capacity as well as ERD comprehension scores before and after the cognitive training. Before the analysis, outliers were eliminated from the data based on the working memory capacity test (AOSpan) threshold. A 2x2 Repeated ANOVA was conducted to compare the effect of (independent variable) *working memory capacity* on (dependent variable) *ERD comprehension scores* in control and experimental groups. The within-subject factors were ERD comprehension score before the treatment and ERD comprehension score after the treatment. The between-subjects factor was the treatment group (control and experiment). There was no statistically significant effect of working memory capacity on the ERD comprehension scores (*Wilk's Lambda*= .969, *F*(1,28)= .885, *p*= .355) (See Table 4). According to this result, we fail to reject the null hypothesis.

Table 4. 2X2 Repeated ANOVA Results

	<i>Wilk's Lambda</i>	<i>df</i>	<i>Error df</i>	<i>F</i>	<i>p</i>	η_p^2
ERD comprehension	.969	1	28	.885	.355	.031
ERD comprehension*treatment group	.950	1	28	1.463	.237	.050

Mean ERD comprehension scores before and after the treatment were 76.842 and 71.228 respectively for the active control group and 70.526 and 71.228 respectively for the experimental group. Although there was no statistically significant difference between control and experimental groups, mean ERD comprehension scores of the active control group decreased after the treatment, whereas mean ERD comprehension scores of the experimental group increased after the treatment.

When descriptive statistics for control and experimental groups on ERD comprehension scores - both before and after the treatment -- were analyzed, an increase in comprehension scores and working memory performance was observed. According to the descriptive statistics, the increase in ERD comprehension scores in the active control group ($M = -5.61$, $SD = 15.85$) was negative, while the experimental group ($M = .70$, $SD = 12.56$) showed increased comprehension scores after the treatment. Similarly, the mean increase in working memory performance for the experimental group was higher. We used two span tasks (AOSpan and Symmetry Span) for working memory performance measure. The Increase in AOSpan was 18% for the experimental group ($M = 17.87$, $SD = 19.76$) and was higher than for the active-control group ($M = 11.20$, $SD = 10.39$). The increase in Symmetry Span performance was 8% in the experimental group ($M = 8.10$, $SD = 23.82$) and higher than for the active control group ($M = -4.29$, $SD = 21.40$).

Although, both groups showed increased working memory performance, mean comprehension was increased only in the experimental group, and decreased in the active control group. To further understand if the mean difference between the groups was statistically significant, a Mann-Whitney U test was conducted. The test indicated that the increase in ERD comprehension scores after the treatment among the participants in the experimental group ($Mdn = -17.57$) was not statistically significantly different from that of the participants in the active control group ($Mdn = 13.43$), $U = 81.5$, $p = .19$. Similarly, the increase in working memory performance after the treatment for the experimental group ($Mdn = 16.50$) was not statistically significantly different from that of the active control group ($Mdn = 14.50$), $U = 97.5$, $p = 0.47$.

5. Discussion

Experiment 1:

To find an empirical answer for the first research question: "How does working memory capacity affect the comprehension skills and eye movements of the novice modelers given a UML activity diagram comprehension task?", we conducted the first experiment and answered the research question as follows:

Results from the first experiment indicate that correctness of the answer given to the ACD comprehension test and working memory performance are important factors that affect reading, understanding, and comprehending the activity diagrams when the modelers are novice, inexperienced undergraduate students.

The findings of the first experiment show that ACD comprehension performance is different for high vs low WM capacity individuals as well as the total number of fixations, average duration of fixations on area of interests, and return time on area of interest. This indicates that novice modelers with high working memory performance are prone to make more fixations than the participants with low working memory performance.

Our study showed that the participants' performance on the ACD comprehension test is higher for those who make more fixations compared to those who make fewer fixations, which is an indicator of an efficient search on complex activity diagram as in line with the previous studies in other domains [27],[28],[29].

This empirical evidence shows that the participants with higher working memory performance focus longer and more frequently on the AOIs in order to extract information compared to the participants with low working memory performance [29], which is also an indicator of the difficulty of activity diagram comprehension test. As we know from the literature, presenting a complex diagram to the participants was important for our study. According to the literature, it is difficult to differentiate the performances when the task is easy. By presenting a complex diagram, the performances of the participants with high and low working memory capacity would differentiate from each other [25]. The number of fixations also led to memory build up [23], which means that each new fixation made by the participant takes up room in the memory

and the greater the number of fixations, the more the memory is built-up. In our experiment, the participants with high working memory capacity were more susceptible to the memory build-up due to the higher number of fixations. Still, they tended to perform better on the ACD comprehension test. In that manner, participants with high working memory capacity were able to effectively process more information in their working memory and tolerate both the high percentage of memory build-up as well as activity diagram complexity better than the participants with low working memory capacity.

An interesting finding is that, compared to those with low working memory capacity, it took longer for the participants with high working memory performance to respond to the activity diagram comprehension question. This metric is also correlated with the correctness of the answer. It took more time to answer the question when the participant's response was correct. According to this finding, working memory capacity can be determinant of persistence in solving challenging questions[3]. Since activity diagrams require high attention and focus to answer the question correctly, especially when the model is complex and the participant is a novice, this finding was expected. The participants wanted to understand the question and extract the information from the diagram carefully by making more frequent and longer fixations on the area of interests correctly and hence, it took longer for the participants with high working memory to answer the question. Yeh and his colleagues [29] found similar results in their study where they focused on visual insight problem solving tasks.

This finding is also supported by the scan path precision and scan path recall values. In our experiment, the ratio of fixations on the area of interest in proportion to all the fixations (scan path precision) was higher for the participants with higher working memory performance compared to the participants with lower working memory performance. Similarly, the percentage of the fixated areas of interests(scan path recall) was higher for the participants with high working memory performance. These findings indicate that our participants with high working memory performed better on the ACD comprehension test, and the way they inspect the activity diagram was more efficient since they focused more on the relevant areas to answer the ACD comprehension questions. This result is in line with the findings in the literature [2].

Specifically, for the ACD comprehension task that we designed, returns to a specific area of interest were crucial for diagnosing the problem and finding a solution to the question. For this purpose, return time was a significant metric for the experiment. Holmqvist and his colleagues [23] suggest that the return time on AOI serves as a measure of working memory. Findings of our experiment illustrate that the return time on AOI for the participants with high working memory capacity was longer when compared to the participants with low working memory capacity. This indicates that the time needed to refresh the information is less for the participants with higher working memory capacity. The results also illustrate that it takes longer for the participants who answered the ACD comprehension questions correctly to revisit the area of interests when compared to the participants who answered the questions incorrectly. This finding shows that the participants with high working memory capacity spend more time processing the information on the area of interests but can process more information in their working memory and, consequently, tend to answer the question correctly. As a result, they do not need to refresh the information for short periods of time unlike the participants with lower working memory capacity.

Experiment 2:

To find an empirical answer for the second research question: "Does working memory capacity training have a positive effect on the entity-relationship diagram understandability and working memory capacity?" we specified the following hypotheses:

- H_0 : There is no statistically significant difference between experimental and control groups in terms of diagram understandability scores.
- H_a : There is statistically significant difference between experimental group and control groups in terms of diagram understandability scores.

Regarding the results of the second experiment we failed to reject the null hypothesis. There is no statistically significant difference in terms of ERD comprehension between participants with high and low working memory capacity. We answered the research question as follows:

The results of the experiment showed that although the difference in ERD comprehension test scores was not statistically significant for active control and experimental groups, the descriptive statistics illustrated that participants in the active control group did not improve their ERD comprehension test scores after the visual search training. On the contrary, their mean ERD comprehension test scores decreased

when compared to their ERD comprehension test scores before the treatment. On the other hand, the experimental group increased their mean ERD comprehension scores after the treatment.

Furthermore, the working memory training slightly improved working memory performance, yet the transfer effect on ERD comprehension task was poor, meaning that improved working memory effect did not seem to improve the ERD comprehension score significantly. The results do not provide empirical evidence that the dual n-back working memory training will significantly improve the ER diagram understandability. Although there is no research in the literature about the effect of working memory training on complex diagram comprehension, this finding is in line with the studies that do not find a transfer effect in other domains [3], [6].

General Discussion:

The findings of both experiments provide empirical evidence that although the working memory training improves the working memory capacity on the training task, it is poor on both improving diagram comprehension and providing a transfer effect. Despite the lack of a statistically significant transfer effect of working memory training on the diagram comprehension tasks, the finding that illustrates slight improvement on the comprehension test score after the dual n-back training. Similarly, a decrease in the comprehension test score after the visual search training shows that working memory training may result in a slight improvement in the diagram comprehension test scores.

Note that this finding could be interpreted to mean that performing challenging working memory training tasks in frequent intervals for 20 sessions over a period of seven weeks was a task that required self-disciplinary behavior and patience on the part of the participants and helped students develop self-disciplinary behavior.

The first experiment showed that high working memory performance leads to persistence in successfully completing the task, paying more attention, and spending more time on the task to succeed. Since the participants of the experiments were novices and the task was challenging, it is understandable that the diagram comprehension scores did not increase substantially between active control and experimental groups. Despite the insignificant difference observed in the second experiment, the eye-tracking data analysis results from the first experiment show that there can be an interaction between working memory capacity and diagram understandability.

6. Limitations of the Study

Certain limitations of the study are discussed below along with procedures we applied to mitigate this study from various threats:

- Lack of a third, no-contact control group mentioned in [30], [31]. The number of participants in our experiments was limited, hence we could not form a no-contact control group. However, literature showed no statistically significant positive impact of working memory training on 3-group experiments: active control; no-contact control; and experimental group [32].
- Internal validity of this research could be threatened by the imprecision of the eye-tracking system. To mitigate this type of risk, we used a large screen during data collection in the first experiment as well as an activity diagram that has sufficient white space around every node. We have also defined the area of interests 0.5 cm larger than the element itself [18]. This could compensate the eye-tracking system's imprecision.
- Another threat to experiments with eye-tracking is fatigue. We mitigated this thread by keeping the tasks short (15 minutes max) and having the participants take a break in between tasks for mental refreshment and ran the calibration again.
- Other studies on model understandability list factors as model-related complexity, color, layout, etc. [33]. We mitigated the effect of model related factors by keeping the diagrams the same for every participant, and the complexity level moderate. There are also personal factors such as modeling expertise, knowledge on notation and diagrams, professional background, domain familiarity, etc. [33]. We mitigated the personal differences by choosing the participants that had similar domain familiarity and diagram familiarity (knowledge) and from the same educational and professional background. We tried to minimize the effect of other factors as much as possible in order to focus on working memory capacity of the individuals.

7. Conclusion

We conducted two different experiments to understand the effect of working memory capacity on diagram understandability. We found evidence that participants' eye movement patterns and diagram comprehension scores differed depending on their working memory capacity: the participants with high working memory capacity tend to perform better, work more efficiently, and persist in finding the

correct answer on the diagram comprehension tests compared to the participants with low working memory capacity. However, working memory training does not provide a transfer effect between working memory tasks and diagram comprehension tasks; nevertheless, we observed a near-significance mean comprehension score increase only in the experimental group as compared to the active control group. Further controlled experiments with a greater number of balanced participants might provide more empirical evidence to support working memory training among novices in complex diagram comprehension.

References

- [1]. Figl, K., & Laue, R. (2015). Influence factors for local comprehensibility of process models. *International Journal of Human-Computer Studies*, 82, 96-110. <https://doi.org/10.1016/j.ijhcs.2015.05.007>
- [2]. Petrusel, R., & Mendling, J. (2013, June). Eye-tracking the factors of process model comprehension tasks. In *International Conference on Advanced Information Systems Engineering* (pp. 224-239). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38709-8_15
- [3]. Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., ... & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2), 359. <https://doi.org/10.1037/a0029082>.
- [4]. Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective?. *Psychological bulletin*, 138(4), 628.
- [5]. Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., ... & Ballard, C. G. (2010). Putting brain training to the test. *Nature*, 465(7299), 775-778.
- [6]. Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829-6833. <https://doi.org/10.1073/pnas.0801268105>
- [7]. Rodrigues, R. D. A., Barros, M. D. O., Revoredo, K., Azevedo, L. G., & Leopold, H. (2015, September). An experiment on process model understandability using textual work instructions and BPMN models. In *2015 29th Brazilian Symposium on Software Engineering* (pp. 41-50). IEEE. <https://doi.org/10.1109/SBES.2015.12>
- [8]. Rodríguez, A., Fernández-Medina, E., Piattini, M., & Trujillo, J. (2008). Secure Business Processes defined through a UML 2.0 extension. *IEEE Latin America Transactions*, 6(4), 339-346. <https://doi.org/10.1109/TLA.2008.4815287>.
- [9]. Dikici, A., Turetken, O., & Demirsors, O. (2018). Factors influencing the understandability of process models: A systematic literature review. *Information and Software Technology*, 93, 112-129.
- [10]. Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- [11]. Gruszka, A., & Necka, E. (2017). Limitations of working memory capacity: The cognitive and social consequences. *European Management Journal*, 35(6), 776-784. <https://doi.org/10.1016/j.emj.2017.07.001>
- [12]. Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic bulletin & review*, 22(2), 366-377. <https://doi.org/10.3758/s13423-014-0699-x>.
- [13]. Au, J., Buschkuhl, M., Duncan, G. J., & Jaeggi, S. M. (2016). There is no convincing evidence that working memory training is NOT effective: A reply to Melby-Lervåg and Hulme (2015). *Psychonomic Bulletin & Review*, 23(1), 331-337. <https://doi.org/10.3758/s13423-015-0967-4>.
- [14]. Kessels, R. P., van Den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment*, 15(4), 426-434. <https://doi.org/10.1177/1073191108315611>
- [15]. Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British journal of psychology*, 95(1), 57-79. <https://doi.org/10.1348/000712604322779460>
- [16]. Đokić, R., Koso-Drljević, M., & Đapo, N. (2018). Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics. *PloS one*, 13(10), e0205169. <https://doi.org/10.1371/journal.pone.0205169>.
- [17]. Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & cognition*, 43(2), 226-236. <https://doi.org/10.3758/s13421-014-0461-7>
- [18]. Petrusel, R., Mendling, J., & Reijers, H. A. (2016). Task-specific visual cues for improving process model understanding. *Information and Software Technology*, 79, 63-78. <https://doi.org/10.1016/j.infsof.2016.07.003>.
- [19]. Sharafi, Z., Soh, Z., & Guhéneuc, Y. G. (2015). A systematic literature review on the usage of eye-tracking in software engineering. *Information and Software Technology*, 67, 79-107. <https://doi.org/10.1016/j.infsof.2015.06.008>
- [20]. Turetken, O., Dikici, A., Vanderfeesten, I. T. P., Rompen, T. M. P., & Demirsors, O. (2020). The influence of using collapsed sub-processes and groups on the understandability of business process models. *Business & Information Systems Engineering*, 62(2), 121-141. <https://doi.org/10.1007/s12599-019-00577-4>
- [21]. Sozen, N. (2020). Experiment Materials. [Google Drive Repository]. Retrieved from: https://drive.google.com/drive/folders/1fbg9_LiFouqY9tGU50nbnrIxxH9jIXtWe?usp=sharing [accessed: 01 August 2020].

- [22]. IQMindware. (2016). Dual N-Back Pro. [Mobile application software]. Retrieved from: <https://apps.apple.com/us/app/dual-n-back-pro/id1082505146>. [accessed: 15 September 2020].
- [23]. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- [24]. Petrusel, R., Mendling, J., & Reijers, H. A. (2017). How visual cognition influences process model comprehension. *Decision Support Systems*, 96, 1-16. <https://doi.org/10.1016/j.dss.2017.01.005>
- [25]. Recker, J., Reijers, H. A., & van de Wouw, S. G. (2014). Process model comprehension: the effects of cognitive abilities, learning style, and strategy. *Communications of the Association for Information Systems*, 34(9), 199-222.
- [26]. Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of memory and language*, 28(2), 127-154.
- [27]. Cagiltay, N. E., Tokdemir, G., Kilic, O., & Topalli, D. (2013). Performing and analyzing non-formal inspections of entity relationship diagram (ERD). *Journal of Systems and Software*, 86(8), 2184-2195. <https://doi.org/10.1016/j.jss.2013.03.106>
- [28]. Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & cognition*, 29(7), 1000-1009.
- [29]. Yeh, Y. C., Tsai, J. L., Hsu, W. C., & Lin, C. F. (2014). A model of how working memory capacity influences insight problem solving in situations with multiple visual representations: An eye tracking analysis. *Thinking Skills and Creativity*, 13, 153-167. <https://doi.org/10.1016/j.tsc.2014.04.003>.
- [30]. Shipstead, Z., REDICK, T. S., & ENGLE, R. W. (2010). Does working memory training generalize?. *Psychologica Belgica*, 50(3), 245-276. <https://doi.org/10.5334/pb-50-3-4-245>
- [31]. Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research.
- [32]. Redick, T., Shipstead, Z., Harrison, T., Hicks, K., Fried, D., Hambrick, D., ... & Engle, R. (2013). No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study. *Journal of Experimental Psychology: General*, 142(2), 359-379. <https://doi.org/10.1037/a0029082>
- [33]. Dikici, A., Turetken, O., & Demirors, O. (2018). Factors influencing the understandability of process models: A systematic literature review. *Information and Software Technology*, 93, 112-129. <https://doi.org/10.1016/j.infsof.2017.09.001>.