Investigating the Influence of Structure on User Performance with UML Interaction Diagrams

Maria Kutar, Carol Britton, Trevor Barker and Jeni Swan
University of Hertfordshire, UK
M.S.Kutar, C.Britton, T.1.Barker, J.Swan@herts.ac.uk

Keywords:
Specification Languages, UML Diagrams, Structure, Understanding of Notations, Empirical Study

Abstract
The importance of structure in specifications and programs has long been recognised in both theoretical research and empirical studies, particularly with reference to the ease with which a reader can understand information representations. In this paper we report on an investigation, using diagrams from the Unified Modelling Language (UML), into whether different structures applied to the same information affect the ease with which readers can understand the information. Our hypothesis was that one of the structure types would produce diagrams that were easier for readers to understand than the other. However, although three studies were carried out, results in each case showed that participants performed equally well on diagrams with each type of structure. We conclude that the difference between two types of structure appears less important than the fact that the two diagram types provide both visible structure and abstractions that the reader may reason with.

1. Background: The Role of Structure in Understanding Diagrammatic Representations

It is generally agreed by authors from both Computer Science and Cognitive Psychology that a clear, easily visible structure is an essential component of an intelligible specification or programme, and the ability to support such a structure is an important property of specification and programming languages (see, for example, Sengler, 1983; Larkin and Simon, 1987; Green, 1989; Winn, 1993; Britton and Jones, 1999).

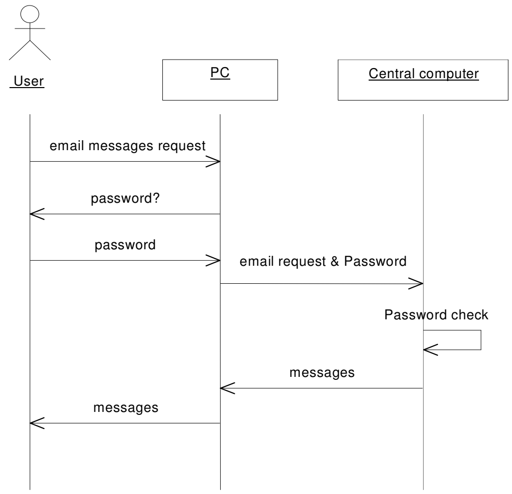
The principal role of a specification is to represent a problem and its solution at various stages of development. If such a representation is clearly structured, it will involve less effort on the part of readers to find, decompose and abstract information, and thus be easier to understand (Green, 1983; Sengler, 1983; Eysenck and Keane, 1990; Winn, 1993). According to Sengler (1983) reading representations of any size involves the activities of decomposition (to split the representation into manageable chunks) and abstraction (to identify the most important features). Decomposition and abstraction are important because a reader can only cope with a small amount of information at a given time. An effective specification language should provide structuring mechanisms to encourage decomposition of the representation into 'brain-sized chunks', each of which is intellectually manageable by the reader. Abstraction helps readers of representations to concentrate on the most important elements while ignoring details that are currently irrelevant. It is important for a specification language to provide a clear, easily visible structure for representations, since, in the absence of a given structure, readers will have to waste time and effort in constructing one for themselves. Any structure is better than none, since it helps the reader to ‘chunk’ the information presented.

Over and above this, a structure will be particularly effective if it in some way reflects the structure of information in the domain, or if it provides the reader with useful abstractions to reason with (Stenning and Oberlander, 1995). Finding, decomposing and abstracting information have been identified as key activities in reading a representation (Larkin & Simon, 1987, Sengler, 1983). Finding information involves searching the representation; this process will be easier to carry out if the language used encourages a clear structure in the representation. A language which provides effective structuring mechanisms is one which encourages developers to produce representations for which the reader does not have to expend effort in working out how to abstract the most important information from it. If the structure of decomposition is not clear, then the reader will be forced to devise his or her own decomposition of the representation, which may well be different from that which the developer had in mind. This again relates to a point made by Stenning and Oberlander (1995) about the importance of ‘specificity’, or clarity in abstraction.

Among languages used to specify software, the imposed hierarchical decomposition of data flow diagrams (see, for example, Fertuck, 1992) mean that it is relatively straightforward for readers to find information at different levels and to concentrate on only a small part of the representation at any given time. Mathematical languages, such as Z (Spivey, 1995), often provide structure at low levels - for example, separation of the specification into schemas can help to identify normal and error cases of an operation - but no mechanism is provided to give an overview of the system as a whole. This adds to the difficulties of readers who are unfamiliar with Z who wish to get a feel for the overall system before examining the details of the representation. An experiment has been carried out to ascertain the implications of imposing a different type of structure on a Z specification; this is reported in (Britton, Jones & Lam, 1998).

Although diagrammatic languages are often thought to encourage more explicitly structured representations then text-based languages, this is not always the case. In an experiment comparing textual and visual programming languages, Green and Blackwell (1996) found that readers' understanding of programs depended more on the structure of the information in the program than on whether the program was written in a visual or text-based language. Green (1983) also makes the point that the structure is only useful if it is clearly visible. Experiments have shown that simple structures are generally more effective than those that try to be 'natural' and mirror the way in which the reader thinks about the problem.

In this paper we describe an experiment on structure using the Unified Modelling Language (UML) (Booch et al., 1999). UML is an industry standard for the development of object-oriented systems and is widely used by developers in both commercial and academic fields. It provides a number of diagram types to support the specification, design and implementation of software systems. These include sequence and collaboration diagrams (collectively known as interaction diagrams), two techniques that produce diagrams that contain the same information, but structured in different ways. This means that these diagrams make an excellent vehicle for an investigation on the effects of different types of structure on the intelligibility of a representation. Examples of a sequence and a collaboration diagram are shown in Figure 1 below. These represent a simple process for reading email.



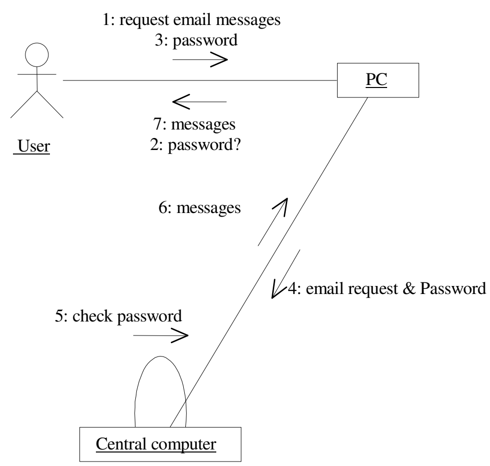


Figure 1: Sequence diagram (top) and collaboration diagram for reading email

The hypothesis in our investigation was that the different structures of the two types of diagrams would mean that one type would be easier than the other for readers to understand. A further hypothesis was that sequence diagrams would be more easily intelligible to readers because their structure more closely reflected the stages in the process represented. In the following sections we describe the three studies which made up the investigation, followed by results and discussion.

2. Study A

2.1 Description

The purpose of the first study described in this paper was to investigate whether users showed greater accuracy in understanding the information contained in sequence or in collaboration diagrams. From the research described above, it was expected that sequence diagrams would result in more accurate user performance. Each of the 124 participants in the study was a first year undergraduate in Computer Science from either the University of Hertfordshire or Anglia Polytechnic University. The experience of the students ranged between having no previous experience with either diagram, to having a little experience of both diagram types. None of the participants claimed to be an expert with either type of diagram.

The study was carried out using a questionnaire, which was produced in four versions and answered anonymously by the participants. The four versions of the questionnaire were distributed randomly amongst the groups of participants. Each version contained six scenarios:

- Making an appointment to see the doctor
- Using a lift
- Driving into a car park
- Ordering a book on the Internet
- Using directory enquiries
- Using a cash machine

In versions 1 and 3 of the questionnaire, scenarios 1, 3, and 5 were represented as sequence diagrams and scenarios 2, 4 and 6 as collaboration diagrams; in versions 2 and 4 of the questionnaire the representations were reversed. In addition, the order of the scenarios in versions 1 and 4 was different from the order in versions 2 and 3.

Each diagram in the questionnaire had five multiple-choice questions relating to the information contained in it; these questions were to be answered by the participants after studying the diagram. Since most of the participants were unfamiliar with the Unified Modelling Language and with these diagrams, the diagrams were referred to in the questionnaire simply as ‘Type 1’ and ‘Type 2’. Participants were asked to state which of the diagram types they thought they would prefer to work with before answering the questions, and which they actually found easier to work with after answering the questions. They were also asked to rate their degree of familiarity with each scenario. At the end of the questionnaire session (which lasted approximately 30 minutes) the scripts were collected and marked; scores were collated and subjected to analysis.

2.2 Results

Table 1 overleaf presents a summary of the data obtained in the study.

Table 1
Summary of the mean data obtained in the study (N=124)

Condition	Mean Score (SD)
Type 1 Diagrams	3.16 (0.1)
Type 2 Diagrams	3.18 (0.1)
Scenario 1	3.12 (0.1)
Type 1	3.03 (0.1)
Type 2	3.21 (0.1)
Scenario 2	3.67 (0.1)
Type 1	3.79 (0.1)
Type 2	3.54 (0.1)
Scenario 3	1.95 (0.1)
Type 1	1.85 (0.1)
Type 2	2.06 (0.1)
Scenario 4	3.73 (0.1)
Type 1	3.67 (0.1)
Type 2	3.77 (0.1)
Scenario 5	2.96 (0.1)
Type 1	2.87 (0.1)
Type 2	3.05 (0.1)
Scenario 6	3.58 (0.1)
Type 1	3.71 (0.1)
Type 2	3.44 (0.1)

An Analysis of Variance (ANOVA) for the data presented in table 1 was performed to determine the significance of any differences in the means shown there. The results of this analysis are shown in table 2.

Table 2

Analysis of Variance performed on the data shown in table 1

Source	df	F	Signif.
Scenario	5	53.01	<0.001
Type	1	0.125	0.72
Scenario * Type	5	1.51	0.19

Table 2 shows a significant difference in mean scores ($p<0.001$) due to the effect of the scenario on performance in the test. There was no significant difference observed due to the effect of diagram type ($p=0.72$).

In order to investigate the effect of scenario familiarity on the scores obtained, scenarios were ranked according to mean familiarity score (1 = most familiar, 6 =

least) and compared with the mean ranked scores obtained by students in each of the scenario This is shown in table 3 below.

Table 3
Mean rank scores obtained by student sand familiarity

Scenario	Rank familiarity 1 = most familiar	Rank Score 1 = highest (easiest)
ATM Machine	1	3
Using a Lift	2	2
Car Park	3	6
Doctor's Appoint	4	4
Dir Enquiries	5	5
Buying a book on Internet	6	1

A Spearman's rank Order correlation was carried out in order to test the significance of any supposed relationship between familiarity and performance. The results of this analysis are shown in table 4 below.

Table 4

Table 4

Spearman's Rank Order Correlation on the data summarised in table 3

N	Spearman's Rho	Significance (Two tailed)
6	-.086	0.87

The results of this analysis show that there was no relationship between familiarity and score ($p>0.05$). It is interesting to note that 'buying a book on the internet' was ranked as the least familiar by participants, yet was the easiest, based on performance in the test.

These results of this analysis are interpreted as follows:

- There was no significant difference between user performance on sequence or collaboration diagrams
- There was a significant difference in mean scores due to the effect of the scenario on performance
- There was no significant difference in scores due to the familiarity of the user with the scenarios.

2.3 Discussion

These results are not in accordance with the original hypothesis that participants would find one diagram (probably the sequence diagram) easier to understand. We suggest that the design of the study may have contributed to this in a number of ways. Firstly, there appears to be some difference in the performance of the participants which relates to the particular scenario rather than diagram type. This indicates that some scenarios were “easier” than others.

In the study, the participants were asked to rate their degree of familiarity with each scenario. Analysis of this information shows that there was no relationship between familiarity and score. Surprisingly, however, the scenario which participants scored most highly on was the one that they felt the least familiar with; this was the scenario related to ordering a book from the internet. We suspect that the lack of familiarity may relate to the purchasing of the book, rather than use of the internet. This seemingly contradictory finding, that participants performed best where they were least familiar with the scenario being represented, could be explained in two ways. Firstly, it could be as a result of participants concentrating harder when they were less familiar with the scenario, and consequently performing better. This would be in line with the findings reported in (Purchase 00). An alternative explanation is that where the participants are less familiar with the scenario they are more likely to read the diagram itself, rather than use their prior knowledge to guess what they think is likely to be the answer.

One potential source of difficulty may have arisen from the wording of the answer categories. Three possibilities were offered to participants: that the statement was true, false or “can’t tell”. The option of “can’t tell” was intended to be used where the statement referred to information which was not present in the diagram. However, it is suggested that many participants may have interpreted it as “don’t know”. Although participants had been told that the study was an investigation into the efficacy of the diagrams, rather than the student’s ability to understand them, we suspect that many students felt that an answer of “can’t tell” would reflect on their competence in using the diagrams. Rather than risking ‘failure’, participants may have guessed at what they think is true or false in the general situation, instead of interpreting the diagram.

In order to eliminate the influence of these factors, a second empirical study was set up. As with the first study, the aim was to investigate the ability of readers to extract information from sequence and collaboration diagrams, and our original hypothesis remained: that the structure of sequence diagrams should result in better performance. We discuss the second study fully in the following section.

3. Study B

One of our primary aims, during the design of the second empirical study, was to eliminate those problems that we had identified with the first empirical study.

3.1 Description

The study was carried out under examination conditions. Participants were informed in advance that some material from their exam scripts may be used for research purposes, and there was opportunity for students to request that their script was not used for this purpose. The participants were second year undergraduate students at the University of Hertfordshire. All participants had attended a course which included coverage of UML sequence and collaboration diagrams. Therefore our participants all had some experience of using the diagrams, but none were experts.

Each of the participants was presented with a compulsory question which was divided into two sections (i) and (ii). In section (i) they were shown a sequence diagram, and in section (ii) a collaboration diagram. The scenarios represented by the diagrams were of similar complexity an were drawn from the same case study, *Willowbank Sports Club*. This case study was one which the students had worked with throughout the duration of the course, and so they were all familiar with it. The sequence diagram represented a scenario where a member unsuccessfully attempted to book a court at a specific time. It is shown in Figure 2 below:

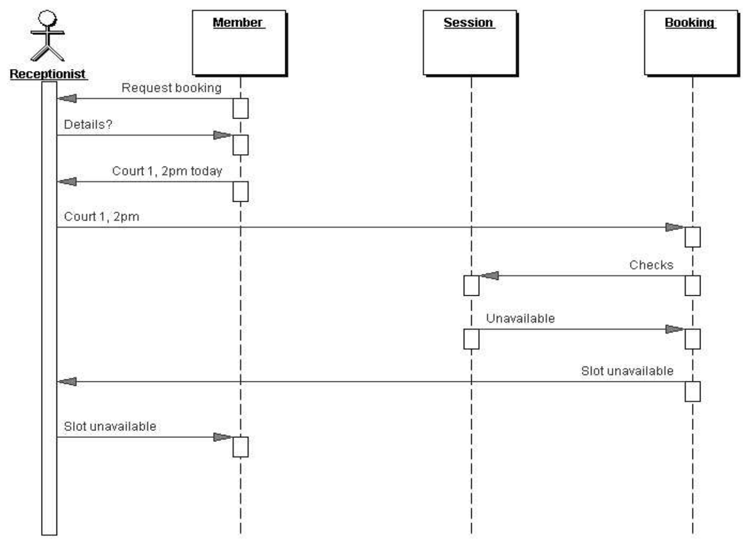


Figure 2: Sequence diagram showing scenario 'unsuccessful booking'

The collaboration diagram showed a scenario where a member cancelled an existing booking (Figure 3 below).

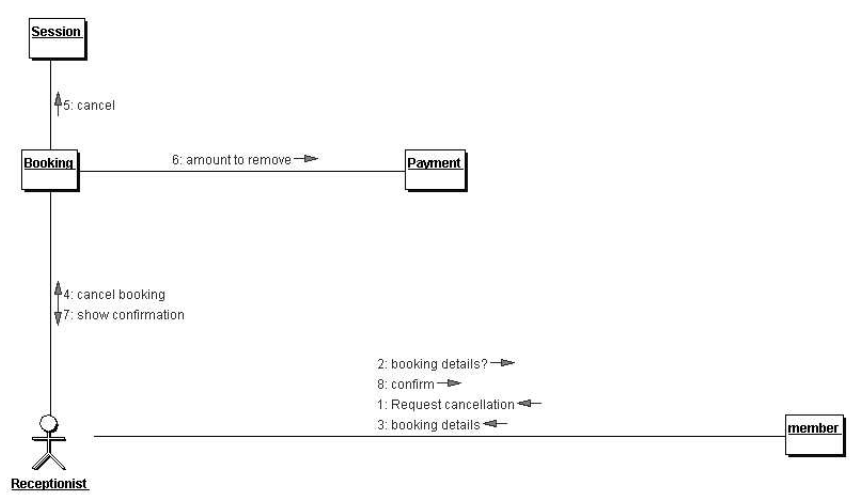


Figure 3: Collaboration diagram showing scenario 'cancel existing booking'

At the start of the question students were given the instruction "Explain in everyday English the scenarios from the Willowbank system that are represented in the diagrams below." This approach, of asking participants to describe the information presented by the diagram in their own words, rather than asking specific questions about that information, was intended to address the problem found in the earlier study relating to the use of “can’t tell” (see section 3.3 above). Such an approach also made clear whether participants were talking about a typical series of events, rather the specific scenario which the diagram represented.

3.2 Results

The mean scores obtained by participants in the second empirical study are shown in table 5 below. The null hypothesis was that any difference between a participant’s scores for diagram types was due to chance alone.

Table 5

Mean scores obtained by participants using sequence and collaboration diagrams

Diagram Type	N	Mean score	Std. Deviation	Std. Error
Sequence	146	2.55	0.562	0.047
Collaboration	146	2.62	0.538	0.045

The data represented in table 5 was subjected to statistical analysis. A paired samples T test was performed to see if the difference in the means for sequence and collaboration diagrams were significant. The results of this analysis are shown in table 6.

Table 6

Paired samples T test analysis of data represented in table 5.

Mean difference	Std. Deviation	Std. Error	T	df	Sig. (2 tailed)
-0.07	0.619	0.051	-1.260	145	0.206

A Pearson product moment correlation was also performed to investigate the significance of any supposed relationship between scores obtained by a participant in part one and part two of the test. The null hypothesis was that any relationship between a participant’s scores for different diagram types was due to chance alone. The results of this analysis are shown in table 7 below.

Table 7

Pearson Product Moment Correlation on the data summarised in table 5

N	Pearson’s Coefficient	Significance
146	0.367	0.001

The difference test shown in table 7 showed that any differences in the mean scores for sequence and collaboration diagrams were due to chance alone ($p>0.05$). The null hypothesis was accepted and there was no difference in performance due to the effect of diagram type. The results of the Pearson correlation show that there was a significant relationship between a subject's score in each diagram type ($p<0.001$) and the null hypothesis was rejected.

The results of this study therefore suggest that the best predictor of performance for a participant when using sequence diagrams is their score for collaboration diagrams, and vice versa. If a subject performs well at one type of diagram, then they perform well using the other type as well. Those performing badly on one diagram type perform badly on both. Any influence of diagram type appeared to be less important in this study.

3.3 Discussion

As with the first study, these results were not in line with our original hypothesis that sequence diagrams would be more readily understood that collaboration diagrams. Whilst we cannot discount the possibility that there is genuinely no difference in user performance with the diagrams despite their differing structure, we must also consider carefully whether the design of the study may have influenced the results.

When designing Study B, we were trying to overcome two identified issues arising from the design of Study A. The first of these was that the wording of the choices for the answers, one of which was "can't tell", may have influenced the results if it was interpreted by participants as "don't know" rather than "information not present in the diagram". We attempted to overcome this by asking students to describe the scenario being represented by the diagram, rather than answering specific questions about it. The second issue was that in Study A there was a significant difference in mean scores due to the effect of the scenario on performance. We identified the degree of familiarity with the scenario as a likely contributing factor to this finding. In Study B the scenarios used were ones of equal familiarity to the participants.

Three particular issues arise from the design of Study B. The first is the difficulty of ensuring that scores were allocated consistently, as differences in participants' descriptions of the scenarios can be attributed to both their use of language and their interpretation of the diagrams. Secondly, the participants' familiarity with the scenarios appears to have influenced their interpretation of the diagrams. For example, some participants described events not shown in the diagram, but which they might reasonably expect to be present in a general case. It is clear that an abstract scenario would be more appropriate for such a study. Finally, as Study B was administered under examination conditions, all participants received the diagrams in the same order (although we cannot be certain that they attempted them in the same order). Therefore an ordering effect cannot be discounted.

A third study, Study C was designed in order to overcome these issues. In this study we returned to the "questionnaire" design, with participants answering questions about the information contained in the diagram. The scenario was designed to force the reader to read the information contained within the diagram itself. We describe this final study in detail in the following section.

4. Study C

4.1 Description

Each of the 133 participants was a second year student at the University of Hertfordshire. They had previously attended a lecture which provided a basic introduction to interaction diagrams, but none of the participants had more than a little experience with either type of diagram.

The study was carried out using a questionnaire, which was produced in four versions and answered anonymously by the participants under supervised conditions. The versions of the questionnaire were distributed randomly amongst the participants. Each version contained one scenario, which showed the operation of a lift. In versions S1 and S2 of the questionnaire, the scenario was shown as a sequence diagram, and in versions C1 and C2 the scenario was shown as a collaboration diagram. The questions relating to information contained within the diagrams were the same in all versions, but were ordered differently in versions S1/S2 and C1/C2.

The scenario, represented as a sequence diagram, is shown in figure 4 overleaf. The scenario is one with which we could expect the participants to be familiar, being based around the operation of a lift. However, it is designed in such a way as to represent a specific sequence of events in the operation of the lift, describing its movement between different floors. Therefore, participants' knowledge of the operation of the lift would not be of assistance to them in answering the question which related to the lift's specific movements as described in the scenario.

Participants were asked to answer twenty multiple choice questions relating to the information contained within the diagrams, choosing between "true", "false", and "information not present in diagram". Example questions are shown below:

- The first thing that happened when the lift reached the ground floor was that the doors opened
- The user got out at the ground floor
- The lift went from floor 45 to floor 28, then to floor 36 and then to the ground floor

4.2 Results

Table 8

Table of mean scores for collaboration and sequence diagrams

Independent variable	**N**	**Mean Score (SD)**
Sequence Diagram	70	13.54 (3.46)
Collaboration Diagram	63	13.13 (3.32)

The data presented in table 8 was analysed using an ANOVA to establish whether any differences in mean scores shown above were due to the effect of the independent

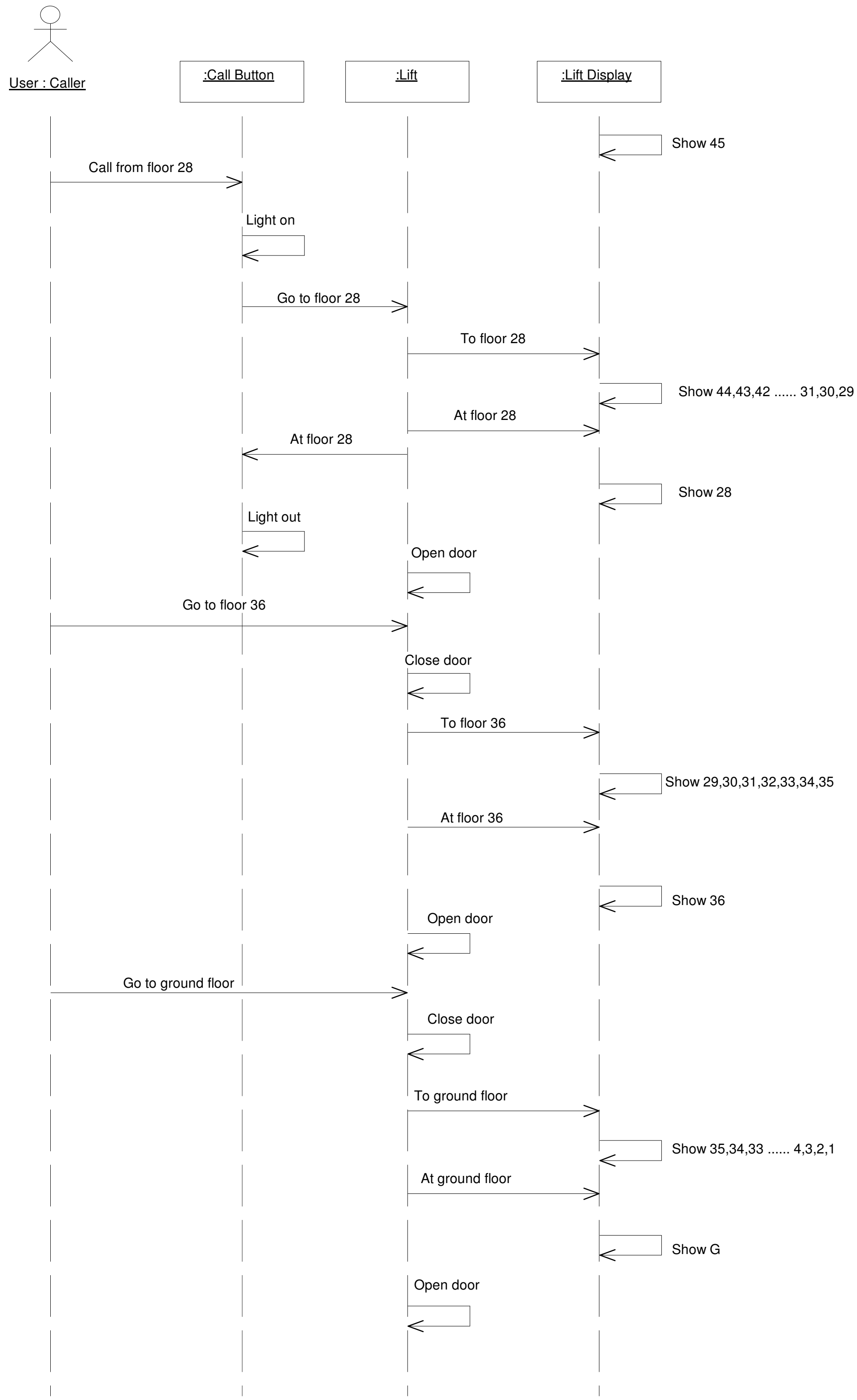


Figure 4: Sequence diagram showing operation of a lift

variable (diagram type), or due to chance alone. The effects of presentation order, age and gender on the score were also analysed. The results of this analysis is shown below in table 9.

Table 9

Results of ANOVA performed on the data shown in table 1

Independent variable	Mean Square	F	Sig
Diagram Type	5.74	0.498	0.48
Order of presentation	2.47	0.25	0.78
Gender	13.45	1.38	0.24
Age	12.38	1.27	0.29

No significant differences were found in any of the conditions. The results of this analysis therefore suggest that the best explanation for any differences in performance on the test were due to chance alone. No significant interactions between age, gender and order of presentation on diagram type were found.

4.3 Discussion

As with the previous studies, the results of this study were not in line with our hypothesis that participants would find one diagram (probably the sequence diagram) easier to understand than the other. In the following section we consider results from the three studies together, and relate them to our initial investigation of the literature.

5 Comparison of Findings and Conclusions

Authors in both Cognitive Science and Psychology are agreed on the importance of structure in a representation and the provision of structuring mechanisms in specification and programming languages. The investigation reported in this paper explored the effect of different types of structure on representations of the same information. Our hypothesis was that the difference in structure would have sufficient impact that participants would perform better with one type of diagram than another. Since the structure of sequence diagrams more closely mirrors the events represented, we felt that it was likely that participants would find these diagrams more readily intelligible.

Three different studies were carried out; each succeeding study attempted to address potential problems in the design of the previous study, for example, the scenario effect, familiarity with scenarios, wording of questions. However, the results from all of the studies indicated clearly that there was no significant difference in the performance of the participants' in relation to diagram type. Although structure may be an important factor in the intelligibility of representations, it appears from this research that the format of that structure may have a much more minor influence. Both sequence diagrams and collaboration diagrams provide structuring mechanisms, and any difference between these mechanisms appears to be insufficient to influence user performance with the diagrams.

Green (1983) makes the point that the structure is only useful if it is clearly visible. This is supported by our studies, as both sequence and collaboration diagrams have a highly visible (although differing) structure. Stenning and Oberlander (1995) claim that a structure will be effective if it provides the reader with useful abstractions to reason with. Both sequence and collaboration diagrams provide useful abstractions: sequence diagrams in terms of ordering of events and collaboration diagrams in terms of links between objects in the domain. In this case the difference between two types of structure appears less important than the fact that the two diagram types provide both visible structure and abstractions that the reader may reason with.

Future work in this area will focus on different representations in order to establish whether differences in structure are sufficient to influence user performance in extracting information from diagrammatic representations.

6 References

Booch, G., Jacobson, I. and Rumbaugh, J. (1999) The Unified Modeling Language User Guide, Addison-Wesley.

Britton, C., Jones, S. & Lam, W. (1998). Separating the system interface from its internal state: an alternative structure for Z specifications. Proceedings of Formal Aspects of the Human Computer Interaction, BCS-FACS Workshop, 87-102, Sheffield Hallam University .

Britton, C. and Jones, S. (1999) The Untrained Eye: How Languages for Software Specification Support Understanding in Untrained Users, Human Computer Interaction, 14, pp. 191-244.

Eysenck, M. & Keane, M. (1990). Cognitive psychology: A student's handbook.. Lawrence Erlbaum Associates.

Fertuck, L. (1992) Systems Analysis and Design, Wm. C. Brown Publishers.

Green, T. (1983) Learning Big and Little Programming Languages, in A. Wilkinson (Ed.), Classroom Computers and Cognitive Science, Academic Press, New York.

Green, T. (1989) Cognitive Dimensions of Notations, in A. Sutcliffe and L. Macaulay (Eds.), People and Computers V, Proceedings of HCI'89, Cambridge University Press.

Green, T. (1991) Describing information artefacts with cognitive dimensions and structure maps, in D. Diaper, and N. Hammond (Eds.), People and Computers VI, Proceedings of HCI'91, Cambridge University Press.

Green, T. and Blackwell, A. (1996) Thinking about visual programs, in Thinking with diagrams, IEE Colloquium Digest No: 96/010, Institute for Electronic Engineers, London.

Larkin, J. and Simon, H. (1987) Why a diagram is (sometimes) worth ten thousand words, Cognitive Science, 11, 65-99.

Purchase, H. C., Colpoys, L., McGill, M., Carrington, D. and Britton, C (2001). UML Class Diagram Variations: An Empirical Study of Comprehension. Proceedings of the Australian Information Visualisation Conference, Sydney.

Sengler, H. (1983) A model of program understanding, in T. Green, S. Payne. and G. van der Veer, (Eds.), The Psychology of Computer Use, Academic Press.

Spivey, J.M. (1989) The Z Notation A reference manual, Prentice Hall.

Stenning, K. and Oberlander, J. (1995) A cognitive theory of graphical and linguistic reasoning: Logic and implementation, Cognitive Science, 19, 97-140.

Winn, W. (1993) An account of how readers search for information in diagrams, Contemporary Educational Psychology, 18.