STRATEGIES FOR USING INSTRUCTIONS IN PROCEDURAL TASKS

A Thesis
Presented to
The Academic Faculty

By

Elsa Eiríksdóttir

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Engineering Psychology

Georgia Institute of Technology

May, 2007

STRATEGIES FOR USING INSTRUCTIONS IN PROCEDURAL TASKS

Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Wendy Rogers
School of Psychology
Georgia Institute of Technology

Dr. Mark Guzdial
College of Computing
Georgia Institute of Technology

Date Approved: March 29, 2007

ACKNOWLEDGMENTS

I want to thank my advisor Dr. Richard Catrambone for invaluable help during all the stages of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The study examined whether an instruction-based strategy (studying the instructions before attempting the task) or a task-based strategy (attempting the task and referencing instructions) was more effective for procedural performance and learning. Four groups of participants learned to perform macramé tasks and assembly tasks, and received detailed instructions at different times in the process of attempting the tasks. Performance was assessed at training and a week later by recording task completion time, correctness, and subjective cognitive load. The strategy for using instructions affected initial performance on the macramé tasks, where instruction-based strategy was superior, but not later retention or transfer. This pattern of results was not found for the assembly tasks indicating that characteristics of the tasks influenced the effectiveness of the strategy for using instructions.

INTRODUCTION

People engage in procedural tasks many times a day, and some of these tasks are well learned and performed effortlessly, such as tying shoelaces or driving. Others are unfamiliar or seldom encountered, like assembling furniture or setting up a stereo system, in which case instructions of some kind are needed to help perform the task. Nearly every product comes with instructions printed on the packages or put in a manual or help systems, whether it is a microwave dinner, a new TV, or a computer game. There are also a wide variety of instructional or how-to books available on everything from building a birdhouse, to cooking Italian food, to doing tai-chi. In addition, the widespread access to the Internet makes instructions for all kinds of procedural tasks readily available.

The most common kind of instructions for procedural tasks, whether printed or electronic, consists of pictures and text explaining each step of the procedure in a linear fashion. Given that instructions are needed and available, how do people make use of them? It is often assumed that instructions should be read before performing a task, but more often than not people do not look at the instructions until they do not know (or cannot guess) what to do next (Ganier, 2004). This study addresses whether the strategy in using instructions influences the learning of a procedural task. In essence, which method of using instructions, reading the instructions before doing the task or using them as reference, is more beneficial to performing and learning a procedure?

Procedural tasks

A *procedure* is doing something by performing step by step actions, and a procedural task is a task that is completed by following those steps. *Procedural knowledge* is knowing the *production rules* that represent the steps needed to perform a task (Newell & Simon, 1972). A common way to describe these rules is to model them as *if – then* pairs, where an action is performed if a condition is met (Anderson, 1993; Koubek, Benysh, & Tang, 1997). A procedural task can therefore be thought of as a series of steps, where each step consists of actions applied if certain conditions are met, according to the production rules.

Each procedural task can be thought to be composed of perceptual, cognitive, and motor components, and these components are differentially important for different tasks. For instance, some procedural tasks lack a perceptual component (e.g., mental arithmetic), others are without a significant motor component (e.g., reading) (Adams, 1987; Ohlsson, 1996). In this study, the focus will be on procedural tasks composed of significant aspect of all three components because it is believed to be representative of procedural tasks in general.

A distinction can be made between two different goals of performing a procedural task. In some cases the goal is simply to perform the task only once, without attempting to learn the procedure. An example of this kind of *one-time procedural performance* is when new furniture is assembled. In other cases the goal is to learn to perform the procedure from memory, and be able to apply it across situations. An example of this kind of *procedural learning* is when children learn to tie their shoelaces. This distinction between one-time procedural performance and procedural learning refers to the goals of the behavior, but the behavior itself might not be different in the two cases. Procedural learning relies on repetition and practice: therefore each time a procedure is repeated it is one-time procedural performance, and some amount of procedural learning presumably takes place during each one-time procedural performance.

The nature of the procedural task is important in terms of how easy or difficult it is to learn and perform. Procedural tasks can be extremely diverse, spanning everything from the very simple (e.g., push a button to turn on a light) to the highly complex (e.g., assembling a car engine). The simplicity or complexity of a task is also contingent upon the materials needed (e.g., tools and parts), and the physical manipulations required for the task. The *affordances*, or the acts permitted by the materials used in the task, are important. For instance, a blank sheet of paper will not have rich affordances for the task of making an origami frog, but a hairdryer with one button has a good indication of its use (Norman, 1988). Most consumers base their decision to read the instructions on whether they think the product needs instructions (Schriver, 1997). This indicates that the affordances will influence whether people use instructions or not; presumably, if a task involves materials with clear affordances, then instructions are less likely to be needed. Alternatively, procedural tasks involving materials with few or no affordances will need instructions. In the same vein, instructions become more important if a task involves unfamiliar or complicated physical manipulations, (e.g., knitting or driving a car with a stick shift) than if the task requires simple or familiar movements (e.g., pressing a button or turning a wheel).

Cognitive factors in procedural learning

Cognitive engagement

An important feature of procedural learning is how much cognitive effort is allocated for learning and understanding the task. Research on procedural tasks indicates that the key to effective procedural learning is to get the learner actively engaged with the task. For instance, Vakil, Hoffman, and Myzliek (1998) showed that active training (learners actively explored the materials) resulted in better performance on a procedural task compared to passive training (following auditory step-by-step instructions), regardless of whether learning performance was measured in terms of using more effective strategy, time to complete the task, or performance on similar tasks. Similarly, work comparing learning for groups that were either presented with a goal-free problem or explicitly stated goal for problem solving, has shown that the goal-free method is more effective (Sweller, van Merrienboer, & Paas, 1998). The goal-free method emphasizes learning by encouraging the learners to explore the problem space, instead of focusing on step-wise problem solving of a particular problem. In addition, guided exploration where instructional materials are designed to specifically encourage active learning has been found to be more effective for learning to use word processing software than a conventional step-wise study manual (Carroll, Mack, Lewis, Grischkowsky, & Robertson, 1985).

There is also evidence that requiring active engagement on behalf of the learner has more effect on learning as measured with problem solving or novel tasks than as measured on isomorphic tasks (identical to the training tasks). Duff and Barnard (1990) found that providing participants with explicit instructions on how to use a computer office system enhanced initial rate of learning as measured by incidence of errors and the time to learn a sequence of actions. This however, led to poorer problem-solving performance on using the system compared to participants who worked out what to do for themselves either unassisted or using general information about the system. These results indicate that even if using detailed instructions can lead to better performance on the training tasks, it might not be as effective when it comes to doing novel tasks. This notion has been supported by work on *self-explaining*, which centers on the idea that good learners study with understanding, that they generate explanations by inference from principles and definitions while examining the study materials. Chi, De Leeuw, Chiu, and LaVancher (1994) demonstrated that learners engaging in self-explanation as they studied task materials performed better on problem-solving tasks than learners who did not. Also, learners who generated a larger number of self-explanation seemed to have a more correct mental model of the subject than those generating fewer self-explanations. In addition, Renkl (1997) showed that quality of self-explanations reliably predicted successful learning from worked-out examples.

Self-explaining can be viewed as elaborate inference, an encoding procedure that generates rich task representations. It is one way of resolving ambiguity associated with study materials, such as examples or instructions, and is invoked when effort is required to work out precisely what to do. However, if instructions provide full specifications on what to do, there might not be any reason for the learner to engage in elaborate inference (Duff & Barnard, 1990; Green, 2002). Green (2002) demonstrated that manipulating task demands and instructions influenced the extent to which learners generated elaborative inferences or engaged in self-explaining activities. In her study Green (2002) varied the amount of what-to-do information given to the participants and found that the group given explicit what-to-do information tended to rely upon a simple rote rehearsal procedure, where apparently little elaboration took place. The group given implicit what-to-do information seemed to be forced to use elaborative inference more frequently to resolve uncertainty about what to do. The conclusion was that the varying the amount of what-to-do information resulted in the construction of qualitatively different task representations.

The commonality of the research discussed is that active engagement with the task and elaborate processing is beneficial to learning, especially when assessed with performance on novel tasks within the domain. Also, there is an indication that step-by-step instructions do not require the learner to engage in active or elaborate processing, but might instead encourage passive imitation and promote the illusion of comprehension (Renkl, 1997). Therefore it is possible that detailed instructions could hinder procedural learning, as the learner relies too much on the instructions and fails to internalize the information. Indeed, Hickman, Rogers and Fisk (in press) showed that learners relying on explicit stepwise instructions in training performed worse when the instructions were removed compared to learners that had training emphasizing implicit guidance of what to do.

Cognitive load theory

Cognitive load theory offers a model of how different cognitive processes involved in a learning activity can either enhance or inhibit learning. The focal point of cognitive load theory (CLT) is how learning can be made more effective by facilitating the cognitive processes involved in learning (Marcus, Cooper, & Sweller, 1996; Paas, Renkl, & Sweller, 2003; Tuovinen & Sweller, 1999). CLT assumes limited cognitive processing capacity, and as more resources are used for irrelevant processing activity, less is available for learning. A central idea is that domain specific knowledge structures, or *schemas*, stored in long term memory allow people to treat many elements as one element. Schema construction and rule automation are the primary functions of learning according to CLT (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Marcus et al., 1996; Paas et al., 2003; Tuovinen & Sweller, 1999).

Because schemas allow for integration of elements, they use fewer cognitive resources and lead to less *cognitive load*. Cognitive load is, of course, partly a function of the cognitive demands of a task; for instance if elements need to be processed individually and considered as discrete elements in working memory, cognitive load will be high (Paas et al., 2003). The sources of cognitive load can be *intrinsic*, *extraneous*, or *germane*. The intrinsic cognitive load of a task refers to the inherent aspects of the mental task that must be understood for the learner to be able to carry out the task. Extraneous cognitive load refers to a range of extraneous factors that interfere with schema acquisition and rule automation; these extraneous factors are usually associated with the way the instructional material is taught. For example, many conventional instructional procedures impose extraneous cognitive load by making the learner search for a solution or referents in an explanation. Germane cognitive load refers to factors inherent in the instructions themselves, the information presented, and the learning activities required of the learner. Germane cognitive load enhances learning because cognitive resources are devoted to schema acquisition and rule automation, but is different from intrinsic load because it is influenced by the instructional design. Intrinsic, extraneous, and germane cognitive loads are additive and the total cognitive load cannot exceed the cognitive resources available if learning is to occur (Paas et al., 2003; Tuovinen & Sweller, 1999).

By having the learner direct more cognitive resources towards schema acquisition and rule automation, learning can be enhanced. One way is to make the learner engage in beneficial learning activities that increase germane cognitive load (Paas, et al., 2003). This could be done by encouraging the learner to actively engage in the task, thereby allowing him or her to build a schema of the procedural task. Renkl (1997) pointed out that self-explanations increase the cognitive load of the learner, and considering their effectiveness, that added cognitive load is presumably germane. Along the same lines, it could be argued that detailed stepwise instructions do not encourage the learner to extend cognitive effort into schema development, as he or she can perform the task by imitation without actively processing the information for understanding.

Strategy for procedural learning

Traditionally, procedural instruction documents are designed to be read before the learner attempts the task and it is assumed that the knowledge is first learned from the instructions and then applied to the task (Carroll et al., 1985; Ganier, 2004; Schriver, 1997). This becomes a linear process because the learner reads through the instructions step by step before applying the information to the task. This linear or *instruction-based* strategy is commonly used by novices or cautious users (Ganier, 2004; Schriver, 1997). Another strategy often seen, but not commonly supported by instructional materials, is more interactive. The learner attempts the task and uses the instructions for referencing. This *task-based* strategy is preferred by experienced users and a subset of novices. Schriver asked users of electronic equipment how they used the instructions for the equipment. She found that 23% of the users read the instructions before they used the equipment, and 42% read the instructions while using the equipment. In addition, 17% of the users referred to the instructions only when they got confused, and 19% did not use the instructions at all. However, because the researchers did not assess prior experience there is no empirical verification that learners with different experience used different strategies. It can be assumed that the users in question had some domain knowledge because the kinds of electronic equipment used in the study, VCR, answering machine, cordless telephone, and a stereo system, are all commonly used. Furthermore, in order to use instructions as references the learner must have some idea about what he or she has to do. Given that learners employ different strategies for using instructions when approaching procedural tasks, which strategy, the task-based or the instruction-based, is more helpful for one-time procedural performance and procedural learning?

At first glance it would seem that learners could clearly benefit from using an instruction-based strategy because it provides them with clear delineation of what to do and how to go about doing it. This might be more beneficial to one-time procedural performance, but as has been discussed, it might actually be detrimental to procedural learning because the learner does not actively engage in what he or she is doing. There are also other reasons for why a task-based strategy might be more beneficial to procedural learning. Alterman, Zito-Wolf, and Carpenter (1991) described a model of instruction usage that emphasized using a task-based strategy. They were concerned with how comprehension guides learning to use a novel device or adjust to novelty in the operation of devices. In their model the operator started by engaging in the activity, not by planning actions from instructions. They argued that instructions are difficult to understand when encountered outside the context of action as they tend to be abbreviated and assume an understanding of the situation. Without context it is unlikely that the learner can comprehend more than just a general sense of what the instructions mean and the operations involved. Also, instructions tend to be phrased in terms of concrete actions and advice, and as such refer to actions that the learner is supposed to be doing.

Fischer and Sugimoto (2006) also emphasized the importance of context when discussing self-directed learning (e.g., learning outside of the classroom). They characterized it as an activity that the learner engages in until a breakdown occurs, when he or she does not know what to do next. For example, breakdown can happen because of incomplete information or because the actions did not have the desired consequences. Breakdown requires the learner to reflect on its causes to find a solution. This is where instructions become helpful by providing support to the learner when he or she needs it and is more likely to actively study them.

The current study

In the current study, I manipulated whether participants engaged in a task-based or an instruction-based strategy by varying when they received detailed instructions for the procedural tasks. Participants learned two different types of procedural tasks; a macramé task where they learned to tie macramé knots, and complete assembly tasks using Zome building blocks. These two types of tasks were chosen because they are presumably unfamiliar to most people and the materials involved do not offer obvious affordances to the learner. They also differ in that the steps in the macramé tasks must be performed in a particular order, but not the steps in the assembly tasks. Thus the macramé tasks can be characterized as more constrained than the assembly tasks. The participants received detailed instructions for the tasks at different times of the learning process; one group of participants studied the instructions before attempting the task (instruction-based strategy), another attempted the task before receiving the instructions (task-based strategy), the third received instructions and attempted the task at the same time, and the fourth did not receive detailed instructions. These last two groups were control groups; the performance of group three was the result of participants using their default strategy and the performance of group four showed level of performance without instructional support. The groups that had to attempt the tasks without detailed instructions were given pictures of the finished tasks for guidance. This was to make sure that they understood what they were expected to do. Performance was measured by the time it took the participants to complete the tasks correctly. One-time procedural performance was assessed at initial training. Procedural learning was assessed in two ways; by having the participants complete the training tasks again a week later (referred to as isomorphic tasks), and by asking them to complete two novel tasks (one on each session), in both cases without detailed instructions. In addition, subjective cognitive load measures were collected on all occasions.

Hypotheses

One-time procedural performance

It was hypothesized that the group using the instruction-based strategy would show better one-time procedural performance than the group using a task-based strategy (Hypothesis 1). This was predicted because working out the procedure without detailed instructions would presumably be more time consuming and error prone than following the stepwise instructions

Procedural learning: Isomorphic tasks

It was difficult to predict results for procedural learning of isomorphic tasks. On the one hand an instruction-based strategy might reduce cognitive load and allow the learner to perform the task correctly and establish appropriate representations of the task from the beginning. On the other hand, a task-based strategy might encourage the participants to become more actively engaged in the task and lead to better learning through more elaborative processing and increased germane load. Therefore, no prediction was made concerning performance on isomorphic tasks.

Procedural learning: Novel tasks

Evidence suggested that the task-based strategy would be more beneficial for procedural learning as measured on performance on novel tasks (Duff & Barnard, 1990; Green, 2002; Vakil et al., 1998). It was therefore predicted that for the novel tasks, a task-based strategy would lead to better performance than instruction-based strategy (Hypothesis 2).

Cognitive load

It was expected that subjective cognitive load would be higher in training for the groups attempting the tasks without instructions. However, it was expected that when performing isomorphic and novel tasks without instructions the groups that received instructions before or simultaneous with attempting the tasks would report higher cognitive load (Hypothesis 3).

Task types

The pattern of results was expected to be the same for both task types, as they both represent procedural tasks, and the study focused on performing and learning procedural tasks in general (Hypothesis 4). However, having information on the effects of strategy for using instructions for two different types of procedural tasks was expected to increase confidence in the generalizability of the results.

Control groups

Furthermore, it was expected that the group receiving instructions and attempting the task simultaneously would generally show the best performance because they would not be constrained to using a particular strategy, but be able to use the instructions and do the task as they preferred. This group was therefore expected to give an indication of performance when participants use their default strategy, that is, use the strategy that they preferred under the circumstances. The participants in the group not receiving instructions were expected to do the worst on procedural performance, representing a baseline of how difficult the tasks themselves are to learn and perform without the aid of any detailed instructions. Comparing procedural learning for the group not receiving instructions and the one receiving the instructions later was expected to indicate whether the detailed instructions hindered or helped learning. Researchers have suggested that giving learners later stepwise instructions, after initial attempts, can possibly discourage them from exerting the mental effort needed for effective learning (Green, 2002). If that is the case, procedural learning should be better for the group not receiving detailed instructions, but if receiving no instructions will hinder learning because the participants have difficulty correcting mistaken assumptions procedural learning should be better for the group getting the instructions later.

METHOD

Participants

A total of 104 participants were recruited from undergraduate psychology courses at the Georgia Institute of Technology. The participants were compensated for their participation with extra course credit. The age range was 18 to 27 years (M = 19.2, SD = 1.5). Participants were randomly assigned to the four conditions, 26 participants in each. Of the 104 participants, 97 returned for the second session. Of the seven that did not return for the second session, six had scheduling conflicts that made it impossible for them to return within the designated time frame (6-8 days after the first session), and one participant chose not to return without giving any specific reason.

Design

A four group (*Inst-Before*, *Inst-Simul*, *Inst-Later*, *No-Inst*) between-subjects experimental design was used. The *Inst-Before* group studied the instructions before receiving the materials for the task. The *Inst-Simul* group received the instructions and the task materials at the same time. The *Inst-Later* group first attempted the task before receiving the instructions. The *No-Inst* group received only the task materials (described below).

Each participant completed two sessions. In the first session (*training session*) participants completed the training tasks using instructions as determined by the condition they were assigned to and then did the first novel tasks without instructions. In the second session (*assessment session*) participants completed the training tasks from the first session without detailed instructions (referred to as *isomorphic* tasks), and then completed the second novel task.

The training session took up to two hours and 30 minutes, while the assessment session was limited to an hour. Performance in terms of time and errors was measured for the four conditions on both sessions, along with subjective measures of cognitive load.

Materials

Macramé tasks

Participants were taught to knot three different macramé knots one at a time; a *Square* knot, a *Vertical* hitch, and a *Pretzel* knot (Lunger, 1998; Meilach, 1971; See Appendix A). The participants were presented with a wooden board with a horizontal bar onto which three pairs of twine threads were fastened, each thread three feet in length. The number of each knot was printed below the bar corresponding to the placement of the threads needed for that knot. The Square knot task (macramé training task #1) consisted of tying a square knot five times; the Vertical hitch task (macramé training task #2) consisted of completing one twist of the tied cord; and the Pretzel knot task (macramé training task #3) consisted of tying three knots in a row. These three types of knots were chosen because they are dissimilar and represented a range of difficulty (Lunger; Meilach). In addition, pilot testing showed that they were simple enough to be done with only a picture of the partially unraveled knot for guidance, but complex enough to make detailed instructions beneficial. The first novel macramé task was a novel way of knotting the square knot (knotting a *Square-twist* knot), and the second was a novel way of knotting the vertical hitch (knotting a *Hitch-chain*).

Zome assembly tasks

Participants were taught to build two different structures using the Zome struts and balls construction tool (see Appendix A). There are four different types of struts that have different colors and differently shaped connectors, and each type comes in three different lengths (short, medium, and long). The struts can be connected using the connector balls; each connector ball has 62 connecting slots: 12 for pentagon shaped struts (red), 20 for triangle shaped struts (yellow), and 30 for rectangle shaped struts (blue). Each participant received a kit with 40 connector balls, 36 yellow struts, 36 red struts, and 56 blue struts. The Zome tool was chosen because it is unfamiliar to most people, has an untypical structure for a construction tool, and it allows one to construct a wide variety of different structures. The structures that were used in the Zome assembly tasks were selected because they represent different level of difficulty, the *Crystal* (Zome training task #1) being the simpler one and the *DNA* (Zome training task #2) more difficult. Only two Zome tasks were used during training (compared to three macramé tasks) because pilot testing showed that they take longer to finish. The first novel Zome task, *Element* was similar in difficulty and structure to the first training task, and the second, *Cube* was of intermediate difficulty. All the structures were designed by the makers of Zome. Pilot testing determined that participants could build these structures using only a colored picture for guidance, but still benefited from using detailed instructions.

Instructions

The detailed instructions were based on published teaching materials for both the macramé task and the Zome task (Lunger, 1998; Meilach, 1971; Zometool, 2006). These instructional materials were used because they represent the type of instructions that are traditionally used for teaching tasks of this kind. The instructions consisted of pictures showing each step in the procedure accompanied by supporting text explaining the step in more detail. Researchers have consistently found that instructions presenting both pictorial and verbal instructions are more helpful for both performance and learning than instructions that use only one medium (Mayer, 2001). Research has also found that learning is better when corresponding text and pictures are presented near rather than far from each other, and simultaneously rather than successively. The instruction design adhered to these guidelines (see Appendix D).

Cognitive Load

An abbreviated computerized version of the NASA Task Load Index (hereafter referred to as NASA-TLX) was used to collect subjective ratings of cognitive workload for each type of training and isomorphic tasks and also for each novel task (Hart & Staveland, 1988) (See Appendix B). This allowed an assessment of the cognitive load inherent in the training and assessment situations, and how difficult the participants found the novel tasks. Collecting subjective ratings of the NASA-TLX with this method has shown high correlations with objective measures of cognitive load (Kalyuga et al., 2001).

Spatial Ability

Both types of procedural tasks used require an understanding of the spatial relations between objects and the manipulation of these objects in three-dimensional space. It was therefore likely that spatial ability, or the ability to generate, maintain, and manipulate mental visual images and figural materials in space, would influence how difficult the learners found the tasks (Ekstrom, French, & Hartman, 1979; Lohman, 1988). The spatial ability of the participants was measured with two tests from the kit of factor-referenced cognitive tests: the cube comparison test and the surface development test (Ekstrom et al.). The cube comparison test measures spatial orientation which has been defined as the perception of position and configuration of objects in space, and the ability of how people perceive spatial relations and orientation with respect to objects in space (Ekstrom et al.). The surface development test measures visualization which is the ability to manipulate or transform the image of spatial patterns (Ekstrom et al.). There has been a debate on whether and how to differentiate between these two spatial ability factors and it has been suggested that spatial orientation refers to perception in space in reference to the observer whereas visualization requires the observer to manipulate the stimulus and alter its image (Ekstrom et al.). Both tests were included in this study because both the ability to perceive objects in space and the ability to manipulate objects in space were relevant to the tasks. The assumption was that together the tests should reasonably indicate the spatial ability of the participants to control for that variable as a confounding factor.

Procedure

Training session: General

Computers were used to control the experiment: this guaranteed that the timing of instruction presentation was accurate and that all the participants received the same mode of presentation and control over the instructions. The participants were randomly assigned to start with either the macramé or the assembly tasks, such that half of the participants in each group started with one task type and half with the other.

After having the experimental procedure explained to them, all participants completed a demographic questionnaire and both spatial ability tests (first the cube comparison test and then the surface development test). Participants were then presented with the task materials and written instructions depending on the condition they were assigned to. There was no absolute time limit for completing the tasks, but because the overall time for the training session was set at two hours and 30 minutes, the experimenter monitored the progress of the participants and moved them to the next task under certain circumstances. If the participant had spent a long time on a single task without any notable progress (for example 20 minutes on a single knot without any results) or showed signs of marked frustration the experimenter asked him or her to move on to the next task. The definition of a long time varied according to task; for the knots that time was 20 minutes per task, for the Crystal structure it was 30 minutes and for the DNA task it was 40 minutes. However, these time limits were not absolute, it in part depended on the progress of the participant; if the participant was about to complete the task, he or she would not be stopped but allowed to complete the task in question. This flexibility was found to be necessary to maximize the use of the time scheduled for the experiment and to minimize participant frustration, but in most cases the participants completed all task within the allocated timeframe without interference.

To guarantee that the participants finished the tasks correctly in the training session, they had to get verification that the task had been completed correctly. Each time the participants believed that they had completed the task they notified the experimenter who then indicated whether the task was correct or not. If the task had been finished correctly the participant could move onto the next task, but if not, he or she had to continue working on the task. Participants never needed more than two verifications, and the experimenter did not give hints as to what needed to be fixed aside from pointing out that the task was incomplete or incorrect. When telling participants that they needed to attempt the task again, the experimenter told them to look more closely at the instructions they were given. This was to give the impression that the experimenter was helping the participant without actually doing so (to minimize participant frustration). The verification process served the purpose of training to criteria, which would allow equal comparison between the groups. It was also meant to collect information on the errors participants made, but as participants never needed more than two verifications, this measure was not very diagnostic.

Once the participants received both the detailed instructions and the task materials, they were able to view the instructions at will, but had to hold down the left mouse button to do so. If the mouse button was not pressed only a picture of the finished task was visible on the screen (see Figure 1). This process provided a way to measure how much time was spent viewing the instructions, as compared to doing the task.

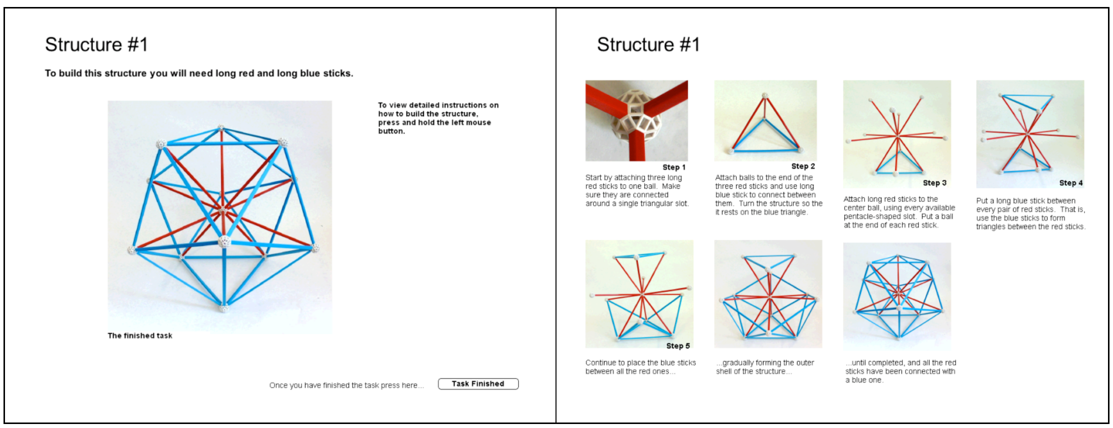


Figure 1. The instructional interface; on the left is the view of the finished task that was seen by all participants. On the right are the detailed instructions were only seen when the left mouse button was pressed (for those participants with access to the detailed instructions).

After completing all the training tasks of one task type (macramé or assembly) and filling out a NASA-TLX measure, the participants completed one novel task for that task type without detailed instructions, after which they completed a NASA-TLX again. To be able to compare performance on novel tasks across sessions, the order of the novel tasks of each task type was counterbalanced. For example, in the training session half of the participants in each group did the Square-twist knot, and the other half did the Hitch-chain, and in the assessment session the participants who did the Square-twist knot in the training session did the Hitch-chain and vice versa.

After completing all tasks of one task type the participants answered some questions about the tasks and instructions; such as whether they liked the task, and whether they thought the instructions were helpful (See Appendix C). To assess prior domain knowledge I asked the participants whether they have ever done anything similar to the tasks before, and if yes, to describe what that was. They were also asked how they normally use instructions to determine what kind of default strategy they normally use when approaching procedural tasks. This information allowed me to determine whether individual differences in strategy had a confounding effect on performance.

Training session: Inst-Before

Participants in the Inst-Before condition were presented with the detailed instructions and asked to study them before attempting the task (see Appendix D). The time for studying the instructions was equal to the time the participants in the Inst-Later group had to attempt the task before receiving the instructions. Pilot testing determined how long the participants had to study the instructions for the tasks, by measuring how much time the participants needed to attempt a task (explained in more detail below). After the designated time (3 minutes and 30 seconds) had elapsed a pop-up window appeared and the participant could start doing the task. The participants had access to the detailed instructions until completing the task.

Training session: Inst-Simul

Participants in the Inst-Simul condition had access to the detailed instruction from the start and could start the task whenever they wanted. They were therefore able to approach the task using their preferred strategy.

Training session: Inst-Later

Participants in the Inst-Later condition were presented with pictures of the completed tasks and asked to start attempting each task using only the pictures for guidance. For the participants to be able to attempt the macramé task without detailed instructions the pictures of the knots showed them both finished and partially unraveled (see Appendix D). This was done because it is impossible to infer the makings of the knots from a picture of them tightly knotted. The pictures used for the assembly tasks showed the finished structure, and were in color to make it easier for the participants to identify which struts were used (see Appendix D). The materials were pilot tested to determine how much time participants needed to attempt the tasks before receiving the instructions. Attempting an assembly task was operationally defined as having tried (successfully or not) to build the structure from at least 20% of the struts needed to complete the task from only a picture of the finished task. Attempting a macramé task was operationally defined as having tried (successfully or not) to tie the knot twice (again using only pictures of the finished task). Pilot data (N = 7) showed that on average participants needed three minutes and 30 seconds to reach the criteria. After the designated time had elapsed, a pop-up window appeared in the middle of the screen notifying the participants that they now had access to detailed instructions.

Training session: No-Inst

Participants in the No-Inst condition were asked to complete the training tasks using only the pictures of the finished task for guidance. Pilot testing showed that this was indeed possible.

Assessment session: All conditions

The second session took place a week after the first session (defined as six to eight days). All participants completed the isomorphic tasks in the same order as in the training session, but without detailed instructions. Participants received only pictures of the finished tasks for guidance (and of the partially unraveled knots in the case of macramé tasks). Half of the participants of each condition received the same order as they did in the first session (e.g. macramé then Zome) and the other half had the opposite order of the first session. This manipulation was to control for order effects. In the assessment session the participant was able to continue onto the next task even if the first was not finished correctly. After finishing the isomorphic tasks for each task type the participants completed one novel task for that task type. After both the isomorphic tasks and the novel task the participants filled out the NASA-TLX questionnaire.

RESULTS AND DISCUSSION

Data preparation

There were 104 participants in the training session (26 per condition), and 97 returned for the assessment session. Three of the participants that did not return were in the No-Inst condition, three were in the Inst-Before condition, and one was in the Inst-Simul condition. Because of time limits in each session (2 hours and 30 minutes for the training session and 1 hour for the assessment session) some participants did not manage to do all the tasks.

To calculate the time-on-task for each training task in the training session, the total duration of the task was converted into seconds, and the duration of viewing the instructions was subtracted. If there was more than one attempt at the task, the time between attempts was also subtracted from the overall duration. For the novel tasks in the training session and all tasks in the assessment session, the duration was simply converted into seconds.

Order effects were tested for task type (whether the macramé or the assembly tasks were completed first) and order between sessions (whether task types were completed in the same order across sessions). No significant effects of order were found ($F < 1$ in all cases), thus order is ignored in the analyses presented below.

Outliers for task duration variables

Outliers were calculated using the time-on-task variable separately for each session and task. This was because the time-on-task was different across sessions for the tasks. The difference was due in part to the manipulations in the training session and in part on differential amount of practice between sessions.

Outliers were defined as data points 3 standard deviations above the mean. If time-on-task for a particular participant was found to be more than 3 standard deviations above the mean, only that data point was removed, other instances of performance by that participant were not removed. The outlier cutoffs were calculated for each task across conditions to reduce standard deviation (R. Ratcliff, personal communication, March 8, 2007). In a comparison of different methods for dealing with outliers, Ratcliff (1993) found that using standard deviation to define outlier cutoffs resulted in good power for ANOVA analysis when the distribution was skewed to the right and when there were large differences in subject mean duration compared to the standard deviation of the distribution. Both these conditions hold for the current study; the duration distribution is skewed to the right for nearly all tasks (see Appendix E) and the difference between largest and smallest participant means (the range) is more than four times the standard deviation for nearly all tasks. The three standard deviations cutoff was selected to conserve potential effects of the experimental manipulation. If the effect of the experimental manipulation is in the tail of the distribution, trimming too many data points can reduce the effect and the power of the analysis (Ratcliff).

Table 1 shows the number of outliers removed from each task and each session with this method. In general, zero to three outliers were removed from each task in each session. In the training session more outliers were removed for the Inst-Later and No-Inst groups than the other two groups. Thus, even using a rather conservative cutoff, more data points were removed in the training session from the two groups expected to have the longest duration. In the assessment session more outliers were removed from the Inst-Before and Inst-Later groups than the other two groups. Altogether 18 data points were defined as outliers in the training session and 10 in the assessment session. Of the 18 data points in the first session 7 were due to the participant struggling with the task and giving up without completing it, and the same was true for 2 of the 10 outliers in the assessment session.

Table 1

Number of outliers removed for each session and task, both overall and for individual condition.

	Training session					Assessment session				
Task	Total	Inst-Before	Inst-Later	Inst-Simul	No-Inst	Total	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Square	3	0	1	0	2	1	1	0	0	0
Vertical	3	0	2	1	0	2	1	1	0	0
Pretzel	3	0	1	0	2	1	0	1	0	0
Crystal	3	0	0	1	2	3	1	1	1	0
DNA	1	0	0	0	1	1	0	1	0	0
Square-twist	1	0	0	1	0	0	0	0	0	0
Hitch-chain	1	1	0	0	0	0	0	0	0	0
Element	2	1	0	0	1	1	1	0	0	0
Cube	1	0	1	0	0	1	0	0	1	0
Total	18	2	5	3	8	10	4	4	2	0

Note: In the training session the total number of participants completing each task was 104, with 26 per condition. In the assessment session the total number of participants completing each task was 97, with 23 in the Inst-Before group, 26 in the Inst-Later group, 25 in the Inst-Simul group, and 23 in the No-Inst group.

One-time procedural performance

Whether participants managed to complete a task correctly was the first indication of how difficult they found the task. The second indication was how many verifications they needed when doing the task. This measure was unfortunately rather limited because participants who completed the tasks correctly did not need more than two verifications. The participants in the Inst-Later and No-Inst groups had a higher percentage of incorrect and incomplete tasks than the Inst-Before and the Inst-Simul groups (see Table 2). The Inst-Later group also had a higher percentage of two-verifications than the other groups, but the No-Instructions group had the lowest.

Table 2
Percentage of correctness and number of verifications for each condition on the training tasks. In parenthesis are the total number of cases and the total possible number of cases; the first number represents the total (100%) in the table, and the second number represents the total cases before outliers were removed (5 training tasks x 26 participants).

	Correctness		Verifications	
Condition	Correct and completed	Incorrect/ Incomplete	One verification	Two verifications
Inst-Before (130/130)	98%	1%	95%	5%
Inst-Later (126/130)	91%	9%	87%	13%
Inst-Simul (128/130)	96%	4%	90%	10%
No-Inst (123/130)	90%	10%	98%	2%

Note: The correctness and verifications are independent. For example, participants not correctly completing the tasks could have either one or two verifications.

Time on task for the training knots was converted to Z-scores to compare performance among the groups on all the macramé training tasks simultaneously. Only cases where the task had been correctly completed were included. After standardization, the results indicated that the participants in the Inst-Before and Inst-Simul conditions were faster than the Inst-Later and No-Inst groups (see Figure 2). A one-way ANOVA revealed a significant difference among the groups, $F(3,285) = 4.47$, $MSE = 0.96$, $p < .01$. A post-hoc Tukey's *HSD* test showed that the difference was significant between the Inst-Later and the Inst-Simul groups ($p < .01$), and marginally significant between the Inst-Later and the Inst-Before groups ($p = .05$). The partial Eta squared (η_p^2) was .05, which means that the instruction manipulation accounted for 5% of the overall (effect and error) variance in time-on-task for the macramé training tasks.

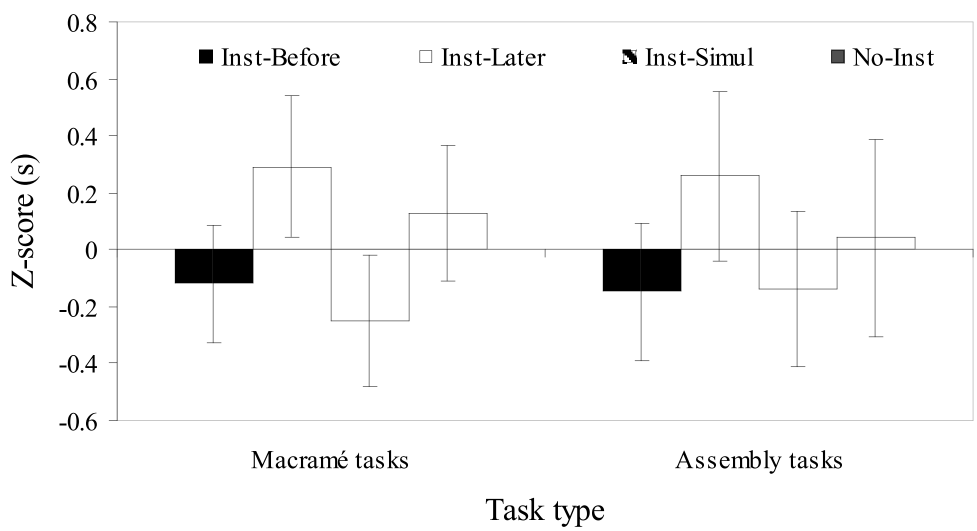


Figure 2. Average standardized time-on-task (in seconds) for the combined training tasks for each task type by condition (zero indicates the mean). The error bars equal plus and minus 2 times the standard error of the mean or the 95% confidence interval.

The results for the assembly training tasks were also combined by standardizing the time-on-task measure. The results had the same general pattern as for the macramé training tasks (see Figure 2). The Inst-Before and Inst-Simul groups were faster than the Inst-Later and No-Inst groups, however the difference among the conditions was not significant, $F(3,185) = 1.78$, $MSE = 0.98$, $p > .05$.

Hypothesis 1 predicted that the Inst-Before group would show better procedural performance than the Inst-Later group. The evidence partly supports the hypothesis. Participants in the Inst-Later group had on average more incomplete or incorrect training tasks, and needed more verifications than the Inst-Before group. There was also a marginally significant difference between the two groups on the macramé training tasks with the Inst-Later group needing more time to complete the task. There was a trend in the same direction for the assembly tasks but the difference was not significant. Hypothesis 4 predicted that the results would be the same for both task types, but the procedural performance results did not support that conclusion. One explanation for the different results for the task types could be that the steps of the macramé tasks need to be completed in a particular order while for the assembly tasks the steps do not have to be done in a particular order. The more constrained nature of the macramé tasks could increase the benefit of having detailed instructions while the unconstrained nature of the assembly tasks could make having the detailed instructions less important.

The Inst-Simul group and the Inst-Before groups had the shortest time-on-task of all the groups for the training tasks and there was not a significant difference between the Inst-Simul and the Inst-Before groups on either task type. This indicates an equal benefit of making participants study the instructions beforehand and having them use their preferred strategy for one-time procedural performance.

It had been expected that the No-Inst group would have the longest time-on-task in the training session, and would be at a disadvantage compared to the Inst-Later group which had access to detailed instructions. The results for the training tasks did not support that prediction, and it indicates that giving participants later stepwise instructions had no benefit over not giving them any detailed instructions at all.

The results for the training tasks therefore suggest an advantage for the Inst-Before and Inst-Simul groups over the Inst-Later group. This advantage was only significant for the macramé tasks, suggesting a differential effect for the two task types. The results for the one-time procedural performance indicate that when completing an unfamiliar procedural task for the first time, people should study the instructions rather than start by attempting the task, especially when the task in question consists of procedural steps that need to be completed in a specific order.

Procedural learning: Isomorphic tasks

In the assessment session the majority of the participants in all conditions completed the isomorphic tasks correctly (see Table 3). The Inst-Before group had the highest percentage of participants that correctly completed the isomorphic tasks. The Inst-Later and the Inst-Simul groups had the highest percentage of participants doing the tasks incorrectly and the No-Inst group had the highest percentage of participants that did not complete the tasks.

Table 3
Percentage of correctness by condition for the isomorphic tasks in the second session. In parenthesis are the total number of cases and the total possible number of cases for each condition.

Condition	Correct and completed	Incorrect/Incomplete
Inst-Before (109/115)	95%	5%
Inst-Later (126/130)	89%	11%
Inst-Simul (124/125)	89%	11%
No-Inst (111/115)	89%	11%

The time-on-task was converted into standardized scores for the isomorphic tasks to allow for comparison among groups for each task type. There was not a significant difference among the groups on the isomorphic macramé tasks, $F(3,250) = 0.72$, $MSE = 0.99$, $p > .05$, or the isomorphic assembly tasks, $F(3,168) = 1.02$, $MSE = 0.99$, $p > .05$ (see Figure 3).

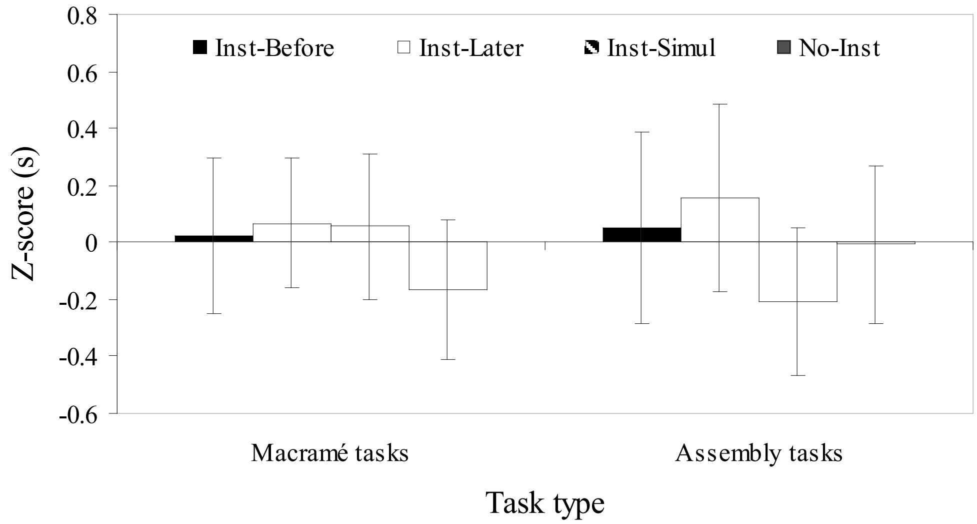


Figure 3. Average standardized time-on-task (in seconds) for the combined isomorphic tasks for each task type by condition (zero indicates the mean). The error bars equal plus and minus 2 times the standard error of the mean or the 95% confidence interval.

There was no prediction made about which strategy (task-based or instruction-based) would be more beneficial for procedural learning of isomorphic tasks. If an advantage had been found for the Inst-Later group (task-based strategy) it would suggest that requiring the learner to become actively engaged in the task before reading the instructions is a better strategy for learning to perform that procedural task compared to following stepwise instructions. However, if an advantage had been found for the Inst-Before group (instruction-based strategy), it would suggest that procedural learning is better when the learner acquires the procedure by studying and following stepwise instructions. The Inst-Before group had a higher percentage of correctly completed isomorphic tasks than the Inst-Later group. However, there was no difference found among the four groups in time-on-task for either task type. It would therefore seem that using different instructional strategy in acquisition does not influence later measures of retention. That is, once a task has been completed, learning has taken place and how the instructions were used ceases to have an effect.

Procedural learning: Novel tasks

The majority of participants did both novel tasks on both sessions, but a few participants ran out of time and had to skip a novel task[1]. Comparison on the novel tasks was made only for participants that completed all training and isomorphic tasks of that task type (macramé or assembly). This was done so equal training and experience with a task type could be assumed for the performance being compared. Table 4 shows the percentage of participants in each group and in each session that completed both the particular novel task in question and all the training and isomorphic tasks of that task type. For example, 81% of all the participants in the Inst-Before group completed all the macramé tasks and the novel knot in the training session. The percentage shown in Table 4 thus represents the proportion of participants in each group that were included in the analysis of the novel tasks.

[1] In the first session, 10 participants did not do one of the novel tasks (two from Inst-Later group, one from the Inst-Simul, and seven in the No-Inst). In the second session 8 participants did not do one of the novel tasks (three from the Inst-Before group, one from the Inst-Later, four from the No-Inst group).

Table 4

Percentage of participants that completed a particular novel task and all tasks of that task type.

	Training session		Assessment session	
Condition	Novel knot	Novel structure	Novel knot	Novel structure
Inst-Before	81%	88%	91%	96%
Inst-Later	96%	92%	96%	92%
Inst-Simul	96%	92%	100%	96%
No-Inst	73%	81%	83%	91%

The correctness of the novel tasks by condition is shown separately for task type (see Table 5). The percentage in Table 5 is of all participants included in the analysis (which is shown in Table 4).

Table 5

Percentage of correctness by condition for the novel tasks of each task type. In parenthesis are the total number of cases and the total possible number of cases for each condition.

Condition	Correct novel macramé tasks	Correct novel assembly tasks
Inst-Before (90/92)	98%	98%
Inst-Later (90/104)	82%	98%
Inst-Simul (93/100)	90%	96%
No-Inst (90/92)	87%	98%

Duration was analyzed for only those cases where participants completed the tasks correctly. To make comparison among the four conditions simultaneously on both novel tasks of each task type across sessions, time-on-task was transformed into Z-scores (see Figure 4). There was not a significant difference among the groups on the novel macramé tasks, $F(3,158) = 0.35$, $MSE = 0.99$, $p > .05$.

The pattern of results turned out differently for the novel assembly tasks. Both the Inst-Before and Inst-Later groups had longer time-on-task than the Inst-Simul and No-Inst groups. A one-way ANOVA showed a significant difference among the four conditions, $F(3,171) = 3.62$, $MSE = 0.94$, $p < .05$.

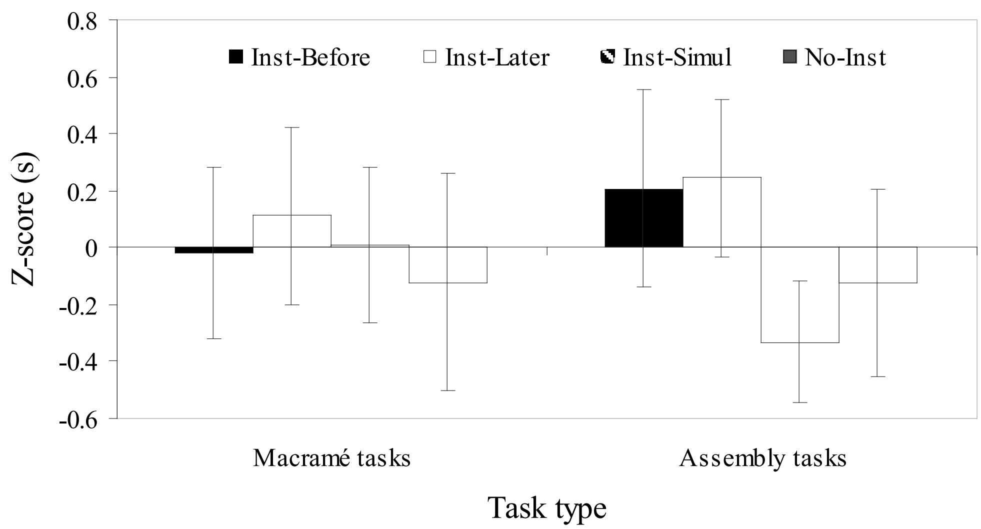


Figure 4. Average standardized time-on-task (in seconds) for the combined novel tasks for each task type by condition (zero indicates the mean). The error bars equal plus and minus 2 times the standard error of the mean or the 95% confidence interval.

A post-hoc analysis (Tukey's *HSD*) revealed this difference to be between the Inst-Simul group and the Inst-Before ($p < .05$) and Inst-Later ($p < .05$) groups respectively. The partial Eta squared was .06, indicating that 6% of the overall variance in time-on-task for the novel assembly tasks was explained by the instruction manipulation.

Hypothesis 2 predicted that procedural learning assessed with performance on novel tasks would be better for the Inst-Later group than the Inst-Before group. The results do not support that hypothesis. The Inst-Before group had a higher percentage of correctly completed novel knot tasks than the Inst-Later group, but there was no difference in correctness between the groups for the novel assembly tasks. In addition, there was not a significant difference between time-on-task for the Inst-Later and Inst-Before groups on the novel tasks for either task type.

No difference was found among the four groups on the novel macramé tasks, but the participants in the Inst-Simul group had a significantly shorter time-on-task (measured in standardized units) for the novel assembly tasks than the Inst-Before and Inst-Later groups.

Therefore, the type of strategy for using the instructions in training did not seem to affect how well the participants were able to transfer the training to novel tasks. The expected advantage of a task-based strategy did not emerge, transfer of training measured by time-on-task showed no benefit of the task-based strategy over the instruction-based strategy. There was an advantage for the participants using their preferred strategy over those made to use either a task-based or instruction-based strategy for the novel assembly tasks. This indicates that participants' preferred strategy resulted in learning that led to better transfer to novel instances of the assembly tasks compared to the other groups. The fact that this was found only for the assembly task could be taken to indicate that the unconstrained nature of the task made it more sensitive to differences in task domain schemas.

Cognitive load

Cognitive load was measured on six different dimensions; mental demand, physical demand, temporal demand, performance, effort, and frustration. In the following analysis the performance dimension is not included. The reason is that all participants consistently reported a high level performance (between 80 and 95 on the scale of 0 – 100; with higher score indicating more success at the task) for all the tasks regardless of condition. This dimension was therefore not considered diagnostic for the cognitive load generated by the tasks in this experiment. For the remaining dimensions higher score indicated more cognitive load, and thus a more negative effect. The scores on the five remaining dimensions were averaged to compose a single cognitive load score.

Cognitive load was measured simultaneously for all the training tasks in the first session and the isomorphic tasks in the second session, but separately for task type in both cases. The analysis was performed only for the cases where tasks had been correctly completed to compare the scores of participants with similar experience of the task (aside from the instruction manipulation); the rationale was that having failed to complete the task could have significant effect on the cognitive load reports.

Because some participants did not complete all the training tasks in the first session or isomorphic tasks in the second session for each task type, the cognitive load measures of all instances of the tasks were aggregated for the analysis (separately for training and isomorphic tasks). For example, if a participant completed all three macramé training tasks, his or her cognitive load scores would be represented three times in the analysis, but if a participant completed only one macramé training task his or her cognitive load scores would only be represented once. Thus, the analysis gave more weight to the scores of the participants who completed all the training or isomorphic tasks.

Training tasks

It was predicted that in training the Inst-Later and No-Inst groups would report more cognitive load than the Inst-Simul and Inst-Before groups and that hypothesis was in most part supported. In the training session the participants in the Inst-Later and No-Inst groups generally reported more cognitive load for the macramé tasks than the other two groups (see Figure 5). For the assembly tasks the Inst-Later group reported the highest cognitive load and Inst-Simul the lowest.

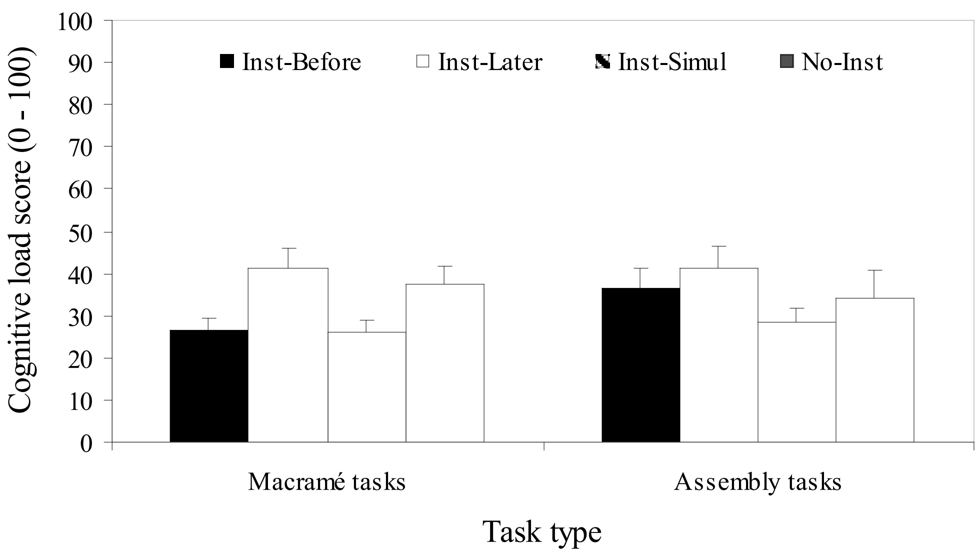


Figure 5. Average cognitive load measures for the training tasks (both task types) by condition. The error bars equal plus and minus 2 times the standard error of the mean or the 95% confidence interval.

There was a significant difference among the groups for both task types (for macramé tasks, $F(3,289) = 18.52$, $MSE = 228.99$, $p < .001$, and for assembly tasks, $F(3,186) = 4.48$, MSE

= 298.34, $p < .01$). The partial Eta squared indicated that the condition explained around 16% of the variance of cognitive load ratings for the macramé training tasks ($\eta_p^2 = .16$), but only about 7% of the variance of cognitive load ratings for the assembly tasks ($\eta_p^2 = .07$). A post-hoc analysis using Tukey's *HSD* test showed that for the macramé training tasks both the Inst-Later and No-Inst groups reported significantly higher cognitive load than the Inst-Before (Inst-Later $p < .001$, No-Inst $p < .001$) and Inst-Simul groups (Inst-Later $p < .001$, No-Inst $p < .001$).
For assembly training tasks the difference was between the Inst-Later and the Inst-Simul group only ($p < .01$).

Isomorphic tasks

It had been predicted that the Inst-Later and No-Inst groups would report less cognitive load than the Inst-Before and Inst-Simul groups for the isomorphic tasks. This prediction did not hold, there was no difference in reported cognitive load for the macramé tasks, and the difference seen for the assembly tasks did not follow the expected pattern.

In the assessment session the cognitive load was generally rated lower than in the training session for both task types (see Figure 6). There was not a significant difference in the ratings among the four groups on the isomorphic macramé tasks, $F(3,254) = 1.36$, $MSE = 258.96$, $p > .05$. For the isomorphic assembly tasks the Inst-Later group reported the highest cognitive load and the No-Inst group the lowest. There was a significant difference among the groups, $F(3,171) = 6.98$, $MSE = 236.05$, $p < .001$.

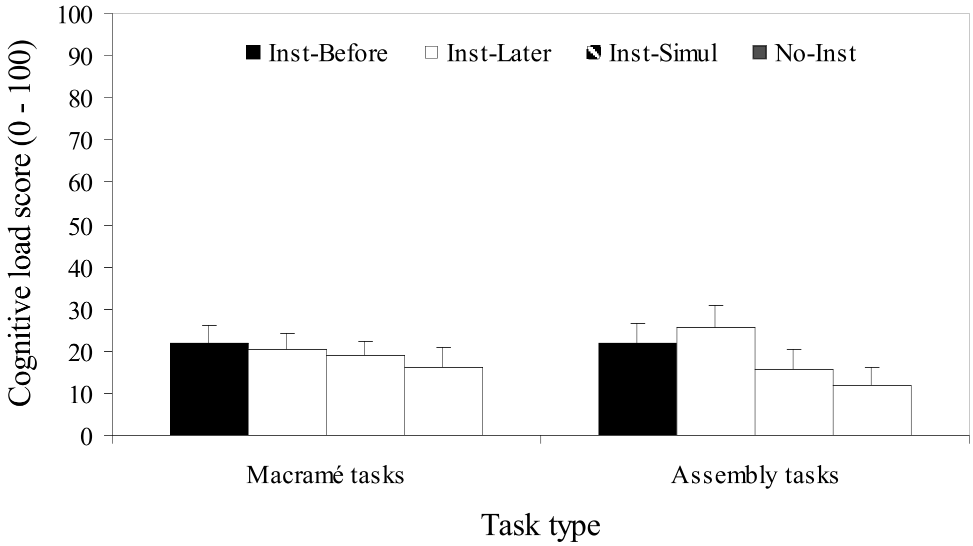


Figure 6. Average cognitive load measures for the isomorphic tasks (both task types) by condition. The error bars equal plus and minus 2 times the standard error of the mean or the 95% confidence interval.

A post-hoc analysis (with Tukey's *HSD* test) showed that the Inst-Later group reported significantly higher cognitive load than the Inst-Simul ($p < .05$) and No-Inst groups ($p < .001$). The Inst-Before group also reported significantly more cognitive load than the No-Inst group ($p < .05$). The condition accounted for 11% of the overall variance in cognitive load for the isomorphic assembly tasks ($\eta_p^2 = .11$).

Novel tasks

For the novel tasks, it had been expected that the Inst-Before and Inst-Simul groups would report higher cognitive load than the Inst-Later and No-Inst groups. The results did not support this prediction; there was no difference in reported cognitive load for the novel macramé tasks and for the assembly tasks the Inst-Later group reported significantly more cognitive load than the Inst-Simul group.

The cognitive load measures for the novel tasks were calculated for each task type and as for the training and isomorphic tasks, the results were analyzed for only those cases where the participants had correctly completed the tasks. For the novel knots the cognitive load ratings were similar across the four conditions, and there was not a significant difference among the groups, $F(3,159) = 1.00$, $MSE = 235.96$, $p > .05$ (see Figure 7). The Inst-Later and Inst-Before groups reported more cognitive load on the novel assembly tasks than the other two groups, and there turned out to be a significant difference among the conditions, $F(3,179) = 4.18$, $MSE = 199.80$, $p < .01$.

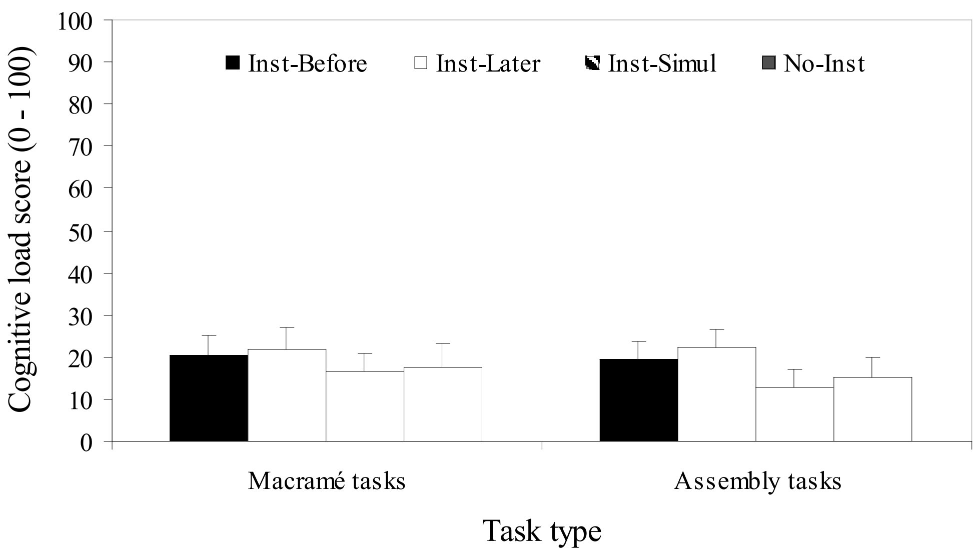


Figure 7. Average cognitive load measures for the novel tasks (both task types) by condition. The error bars equal plus and minus 2 times the standard error of the mean or the 95% confidence interval.

A post-hoc analysis (with Tukey's *HSD* test) showed that the difference was between the Inst-Later and Inst-Simul groups ($p < .01$). The condition accounted for about 7% of the overall variance in cognitive load for the novel assembly tasks ($\eta_p^2 = .07$).

Cognitive load results for the training tasks indicated more cognitive load for the Inst-Later and No-Inst groups than the other two groups for the macramé tasks. There was less difference in cognitive load for the assembly tasks; a significant difference was only found between the Inst-Later and Inst-Simul groups. For the isomorphic tasks, there was no difference in reported cognitive load for the macramé tasks, but the Inst-Later reported more cognitive load on the assembly tasks than the Inst-Simul and No-Inst groups. The results for the novel tasks was similar; no difference in reported cognitive load was found for the novel knot tasks, but the Inst-Later group reported significantly more cognitive load than the Inst-Simul group on the novel assembly tasks. Cognitive load results therefore indicate a difference for the two task types, with a difference among the groups on the macramé tasks at training, but not at transfer or later retention. For the assembly tasks however, the Inst-Later group reported significantly more cognitive load than the Inst-Simul group on all occasions.

It was predicted that the participants in the Inst-Later and No-Inst groups would report higher cognitive load than the participants in the Inst-Simul and Inst-Before groups during training and that hypothesis was mostly supported. It was also predicted that the opposite would hold for performance on isomorphic and novel tasks, but that hypothesis was not supported. These results indicate that the participants with later or no access to detailed instructions experience more cognitive load in training on tasks that are relatively constrained (where steps have to be completed in a specific order) compared to participants that have access to detailed instructions from the start. However, when the tasks are less constrained (steps can be completed in different order) participants with later access to detailed instructions experience more cognitive load than those using their default strategy not only in training, but also in retention and transfer.

Individual differences

The individual difference measures collected were GPA, SAT scores (verbal and math), and two measures of spatial ability (spatial orientation and visualization). The descriptive statistics for each of the measures are given in Table 6.

Table 6
Descriptive statistics for individual difference measures.

Measure	M	SD	Minimum	Maximum
GPA	3.2	0.6	1.4	4.0
SAT-Verbal	635.5	74.6	400.0	800.0
SAT-Math	690.3	63.5	470.0	800.0
Spatial orientation*	20.1	11.3	0.0	42.0
Spatial visualization*	40.7	18.0	0.0	60.0

* The spatial ability tests are scored to minimize guessing, that is, the numbers of wrong answers were subtracted from the number of right ones. This means that there are some participants that ended up with 0 as a score. For the spatial orientation test the maximum possible score was 42, and for the visualization the maximum possible score was 60.

There were no significant differences among the groups on SAT-Verbal score, SAT-Math scores, spatial orientation, or spatial visualization ($F < 1$ in all cases). There was however a marginally significant difference found for the groups on GPA, $F(3,86) = 2.59$, $MSE = 3294.94$, $p = .06$, and post-hoc analysis (Tukey's *HSD* test) showed that this difference was between the Inst-Before and the Inst-Later groups ($p = .07$). The participants in the Inst-Before group had higher GPA ($M = 3.45$, $SD = 0.44$) than the participants in the Inst-Later group ($M = 3.02$, $SD = 74.16$).

On a post-task questionnaire[2] information was collected on whether the participants had ever done anything similar to the two task types (i.e. macramé and assembly) before. It turned out that 64% of the participants had never done anything like the macramé tasks before and 60% had never done anything like the assembly tasks before. Of the participants reporting familiarity with the knots most mentioned friendship bracelets, boy- or girl-scouts, or climbing as the reason for familiarity. Most of the participants reporting familiarity with assembly tasks mentioned Tinker-Toys, Lego, and K'nex as the source of familiarity. For both task types, about 40% of participants in each group reported being familiar with the task, with the exception of the Inst-Later group. Only 22% of the participants in the Inst-Later group reported familiarity with the macramé tasks.

In addition, it was assessed whether there was a match between the self-reported default strategy for using instructions and the condition the participants were assigned to (specifically the Inst-Later and Inst-Before groups). When asked to choose whether they would normally start a procedural task by first studying the instructions or attempting the task, almost 70% of the participants in the Inst-Before and Inst-Later groups said they used a task-based strategy (see Table 7).

[2] For completeness the results of the questionnaires are summarized by condition in Appendix F.

Table 7
Percentage of participants preferring either instruction-based or task-based strategy.

When you do something for the first time do you normally:	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Start by studying the instructions?	32%	31%	52%	64%
Start by attempting the task?	68%	69%	48%	36%

The participants in the Inst-Before group that said they start by studying the instructions were coded as having a strategy match, and the same was done for the participants in the Inst-Later group that said they started by attempting the task.

A step-wise multiple-regression was used to investigate the effects of the individual difference variables on time-on-task. The proposed predictors were GPA, SAT scores, spatial ability scores (spatial orientation and spatial visualization), familiarity with the tasks, age, gender, order of tasks (training and isomorphic), and strategy match. Additionally, condition was added to the analysis. Dummy-coding was used to code the categorical predictors. For example, order of macramé tasks was represented by two variables that would take on values of 0 or 1 (0, 0 representing the first task or Square knot; 0, 1 representing the second task or Vertical knot; 1, 0 representing the third task or Pretzel knot). The regression was calculated separately for the macramé and the assembly tasks. The variables were entered stepwise using the criteria that the predictor with the largest F-to-enter value (calculated at each step) was entered next. The criteria set for *F*-to-enter were at .05 and *F*-to-remove at .1.

Table 8 summarizes the regression for the macramé tasks. The predictors that ended up in the regression equation were the order of knots and familiarity with knots. Participants were faster completing both the second (Vertical) and the third (Pretzel) knot than the first (Square).

Participants familiar with knots were also faster at completing the macramé tasks than the participants not familiar with knots.

Table 8
Summary of stepwise regression for the macramé tasks.

Variable	Final Coefficient	Standard Coefficient	R^2 Total	Correlation
y – intercept	358.12			
Task-order: Vertical	-165.17	-.43	.06	-.253**
Task-order: Pretzel	-134.19	-.35	.16	-.132**
Familiarity with knots	64.83	.17	.19	.175**

*p < .05. **p < .01.

Potential predictor variables that did not end up in the final equations were: condition, age, gender, GPA, SAT scores, spatial ability, and strategy match. Overall, 19% of the variance in all the macramé tasks combined was accounted for by the regression equation.

The results for the assembly tasks were dramatically different. In fact, none of the predictor variables reached criteria for entering the regression equation. This indicated that the variance in time-on-task for all the assembly tasks combined was not due to any of the identified predictor variables: condition, age, gender, GPA, SAT scores, spatial ability, order of tasks, familiarity with tasks, or strategy match.

The regression analysis showed limited influence of the measured individual difference variables on performance. Order of task and familiarity with knots predicted performance on the macramé tasks, but none of the variables entered into the equation predicted performance for the assembly tasks. The results indicate a difference between the two task types, prior familiarity with knots helped performance (in terms of time-on-task) while prior familiarity with assembly tasks did not. Also, having completed one macramé task predicted shorter time-on-task for later tasks, while having completed one assembly task did not predict any change for later task performance. This might be due to the fact that the macramé tasks are in many ways more similar compared to the assembly tasks. That is, all the macramé tasks require that each step is completed in a particular order and then each knot must be completed a number of times, which means that the same steps are carried out in the same order successively. The assembly tasks may appear more diverse, and each task can be tackled by a different method as the steps of the task do not need to follow a certain order. The perceived coherence of the tasks belonging to each task type might therefore help to determine how relevant or helpful prior experience is.

It was expected that spatial ability could predict performance on both task types, but neither spatial ability test was found to be a significant predictor of performance. Scatter plots indicated that restriction of range in the spatial ability tests is not a plausible explanation for this finding. It would therefore seem that performance on these task types is less dependent on spatial ability than originally thought or that the relevant spatial ability dimensions were not adequately captured by the two spatial ability tests. It was also interesting to find that performance was not better for the participants with a match between their preferred strategy and the strategy they were made to use in the study. This can be taken to indicate that the success of making people adopt a new strategy for using instructions is not dependent on the strategy they are accustomed to.

CONCLUSIONS

The results showed that for one-time procedural performance the participants using an instructions based-strategy for the macramé tasks had better performance and reported less cognitive load than the participants using a task-based strategy. This was not found to hold for the assembly tasks, there was no difference in performance or reported cognitive load. There was no difference between participants using either strategy when completing the tasks again a week later. This can be taken to indicate that the strategy for using instructions did not affect later retention of the tasks. Whether the participants used a task-based or instruction-based strategy did not seem to affect transfer of training to novel tasks. Overall the results indicate that using either a task-based or instruction-based strategy in learning a procedural task can affect performance in training, with an advantage for the instruction-based strategy. However, the strategy for using instructions does not seem to be as important for later retention or transfer performance. This last conclusion needs to be tested more thoroughly because the alternative explanation that the method was not sensitive enough to detect the difference cannot be excluded. This indicates that how people use instructions only affects initial performance or acquisition and that other factors become important once basic learning has taken place.

One weakness of the current study is that the variables measured did not allow the identification of what predicted retention and transfer of learning on these tasks. Another problem is that it was difficult verifying the exact strategy each individual participant used in training. Behavioral verification (observing whether tasks were indeed attempted and instructions studied) might not be enough. For example, even if a participant was made to attempt the task before receiving the instructions he or she could still only superficially engage in the task (not making an effort) while waiting for the instructions to become available[3]. Therefore, there is no guarantee that every participant made to use either the task-based or the instruction-based strategy was actually fully committed to doing so. Further research should attempt to get participants motivated to employ a certain strategy and include a subjective report to verify that this strategy was indeed employed.

There were unexpected differences found between the two task types (macramé and assembly), indicating that attributes of the tasks in question influence whether a strategy is beneficial for procedural performance. Further research needs to be done to untangle these issues, for example by identifying the attributes of the tasks that determine which strategy is beneficial. In this study the task differences were roughly defined as the assumption was that the strategy for using instructions would have similar effects across tasks. Future studies should use tasks that can be modified to include different attributes, for instance being either constrained (requiring a particular order of steps) or unconstrained (steps can be performed in any order). Performance could also be measured with more precision by using computerized tasks; one problem with the current study was the difficulty of tracing errors. It was assumed that errors would increase time-on-task, but it would be more informative to be able to trace exactly what errors are made where in the process of completing the task. This can only be done by either videotaping performance or using a computerized task.

There were two main results of the current study. First, the strategy for using instructions seems to influence only initial performance on procedural tasks, and have little or no effect on later retention or transfer. Secondly, characteristics of tasks influence the effectiveness of the strategy for using instructions. The issue of how people use instructions seems therefore to be

[3] All participants in the Inst-Later group attempted the tasks before receiving instructions, and all participants in the Inst-Before group appeared to study the instructions before attempting the tasks.

both more limited in scope than initially believed and also more complex. From the results of this study the recommendation for instructional design would be to require the learner to study the instructions before attempting the task, especially when the steps of the task need to be completed in a specific order.

APPENDIX A

TASKS

Macramé tasks

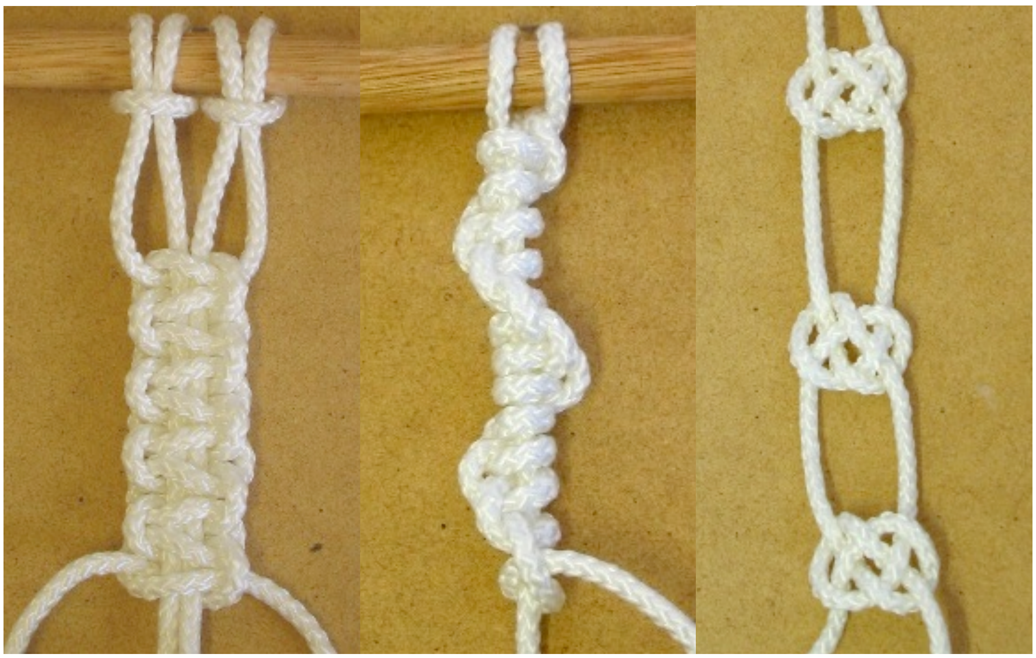


Figure A1. Macramé training/isomorphic tasks; far left is the Square knot, in the middle is the Vertical hitch, and far right is the Pretzel knot.

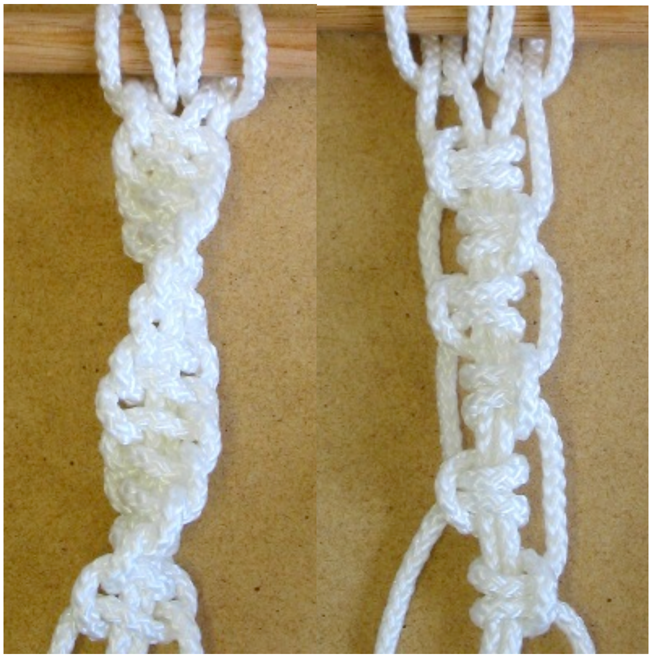


Figure A2. Macramé novel tasks; to the left Square-twist knot and to the right Hitch-chain

Assembly tasks

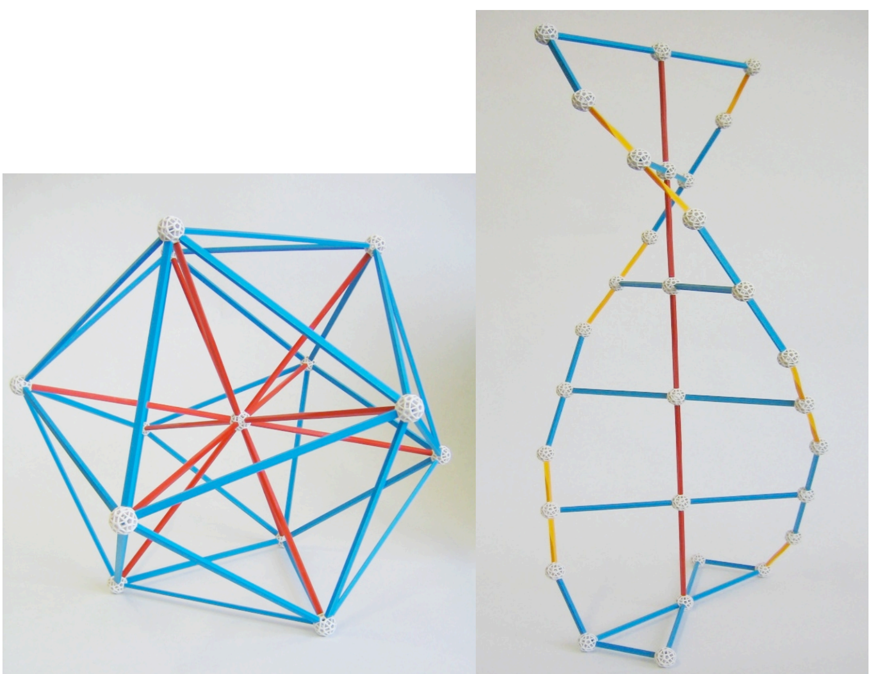


Figure A3. Assembly training/isomorphic tasks; to the left Crystal, and to the right DNA.

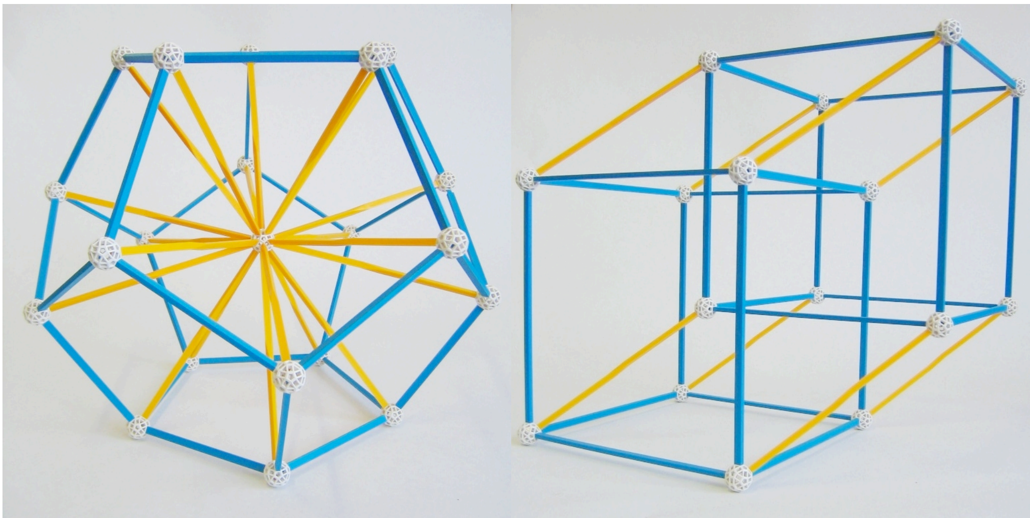


Figure A4. Assembly novel tasks; to the left Element, and to the right Cube.

APPENDIX B

COGNITIVE LOAD MEASURES

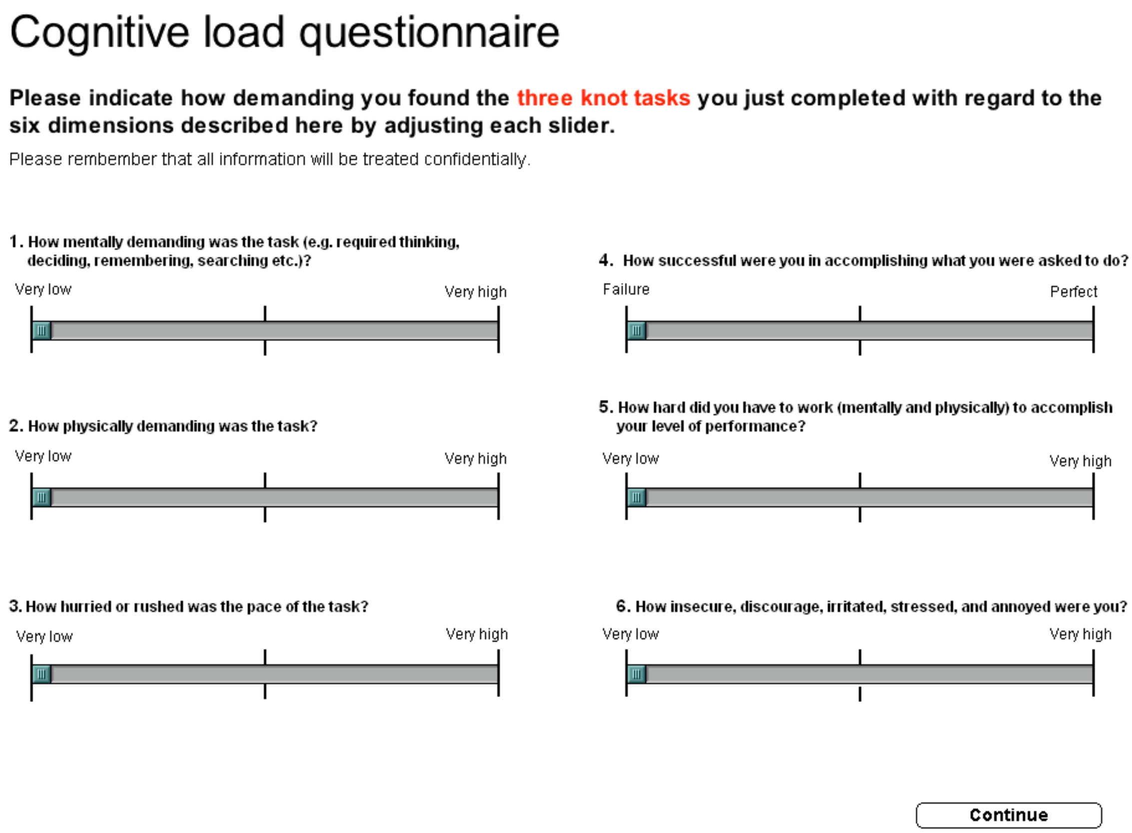


Figure B1. The interface of the abbreviated NASA-TLX to measure subjective cognitive load.

APPENDIX C

QUESTIONNAIRES

Questions about the knots

We would appreciate your opinion on the instructions and the tasks you just finished.

Please remember that all information will be treated confidentially.

1. Have you ever done anything like the knot tasks before? ○ Yes ○ No

If yes, please describe it in a few words [ ]

If yes, did it help you with the current tasks? ○ Not at all ○ Not really ○ Neutral ○ Somewhat ○ Very much

2. Did you like the knot tasks? ○ Not at all ○ Not really ○ Neutral ○ Somewhat ○ Very much

3. Did you find the knot tasks difficult? ○ Not at all ○ Not really ○ Neutral ○ Somewhat ○ Very much

If so, can you describe what it was that you found difficult? [ ]

4. Were the instructions helpful? ○ Not at all ○ Not really ○ Neutral ○ Somewhat ○ Very much

5. Did you find the instructions easy to use? ○ Not at all ○ Not really ○ Neutral ○ Somewhat ○ Very much

6. When using the instructions, which part did you focus the most on?

○ The pictures primarily

○ Mostly the pictures, but skimmed the text

○ The pictures and the text equally

○ Mostly the text, but glanced at the pictures

○ The text primarily

[Continue]

Figure C1. Post-questionnaire for the macramé tasks (identical for the assembly tasks)

Final questions

Please remember that all information will be treated confidentially.

1. How do you usually use instructions? (For example instructions for using, assembling or troubleshooting things)

- ○ I read the instructions thoroughly before beginning the task
- ○ I skim the instructions shortly before beginning the task
- ○ I reference the instructions once in a while to verify what I am doing
- ○ I only use the instructions if I run into problems
- ○ I never use the instructions

2. When you do something for the first time do you normally

- ○ start with studying the instructions?
- ○ start by attempting the task?

3. Please indicate your familiarity with the activities listed below

Using Lego	○ Not familiar	○ Somewhat familiar	○ Very familiar
Using Magnetix	○ Not familiar	○ Somewhat familiar	○ Very familiar
Using Mekano	○ Not familiar	○ Somewhat familiar	○ Very familiar
Making friendship bracelets	○ Not familiar	○ Somewhat familiar	○ Very familiar
Braiding	○ Not familiar	○ Somewhat familiar	○ Very familiar

Continue

Figure C2. Final questionnaire in the training session.

APPENDIX D

EXAMPLES OF INSTRUCTIONS

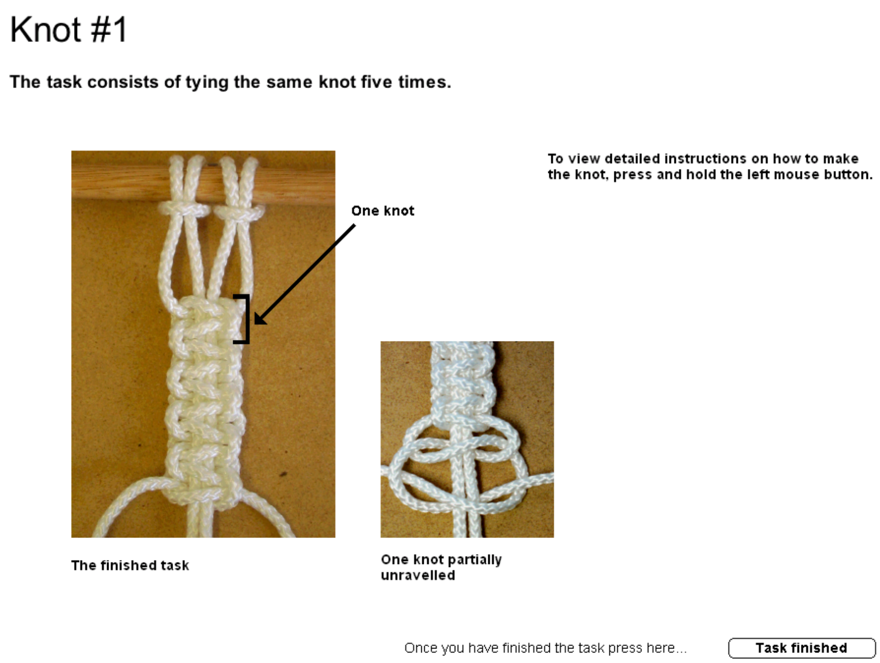


Figure D1. Pictures for guiding the tying of the Square knot without instructions.

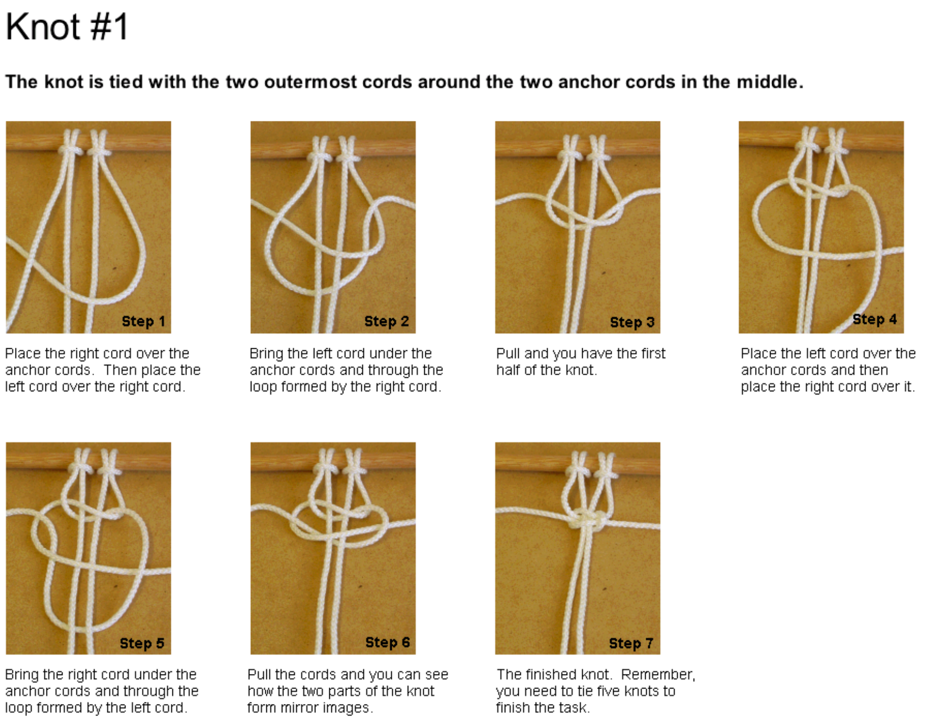
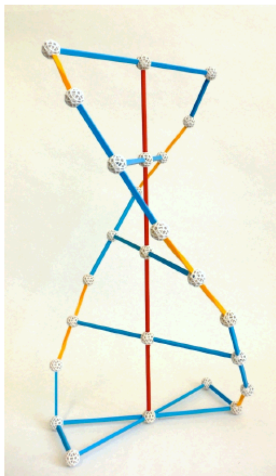


Figure D2. Instructions for the Square knot (accessible by holding the left mouse button).

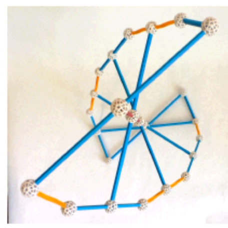
Structure #2

To build this structure you will need short and medium blue sticks, short yellow sticks, and medium red sticks.

To view detailed instructions on how to build the structure, press and hold the left mouse button.



The finished task in sideview



The finished task seen from above

Once you have finished the task press here... **Task Finished**

Figure D3. Pictures for guiding the building of the DNA structure without instructions.

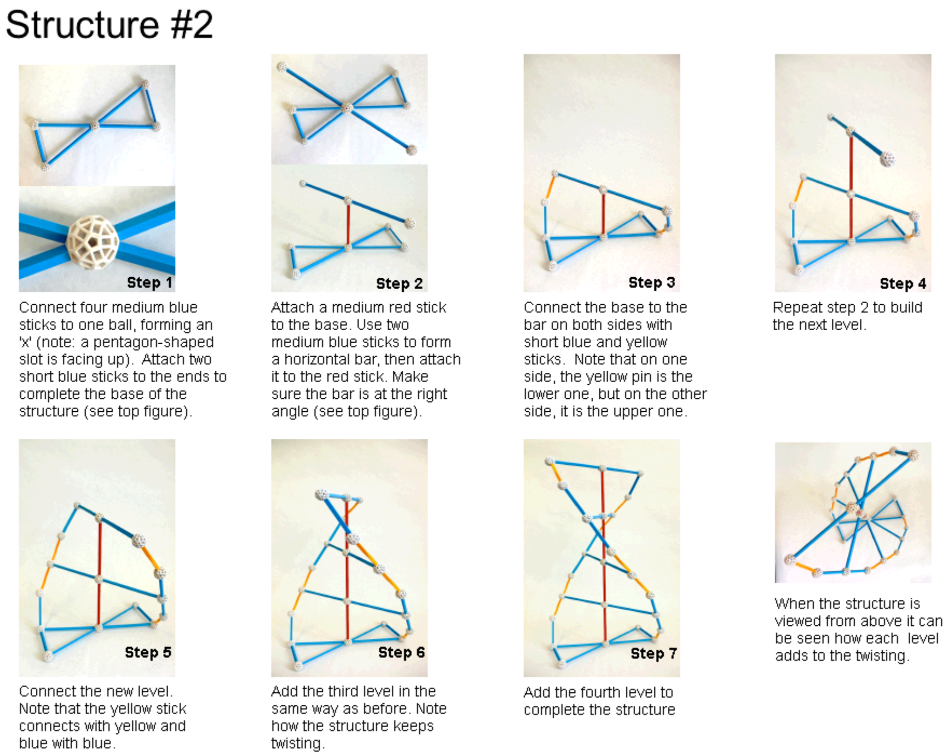


Figure D4. Instructions for the DNA structure task (accessible by holding the left mouse button).

APPENDIX E

DURATION DISTRIBUTIONS AND OUTLIERS

Training tasks (training session)

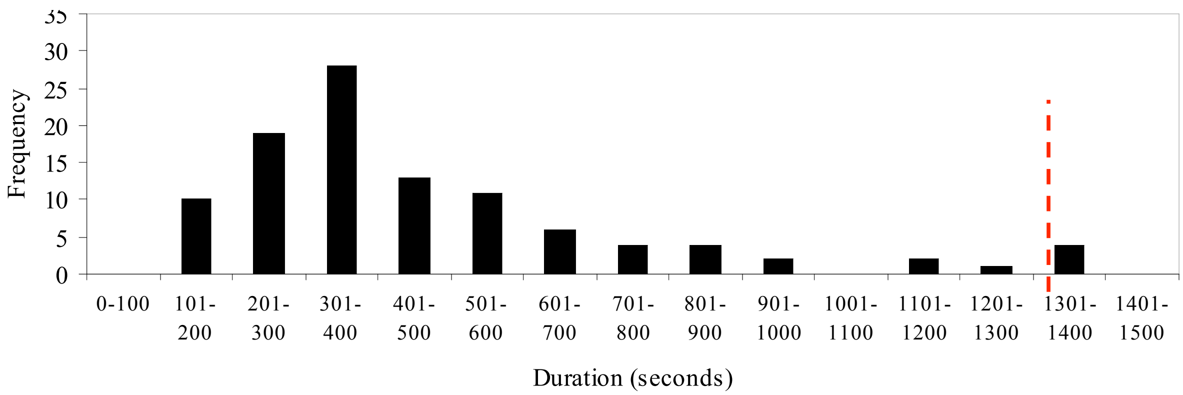


Figure E1. Duration distribution for the Square knot task (*M* = 470.58, *SD* = 284.20). Red dashed line represents the three standard deviation outlier cutoff.

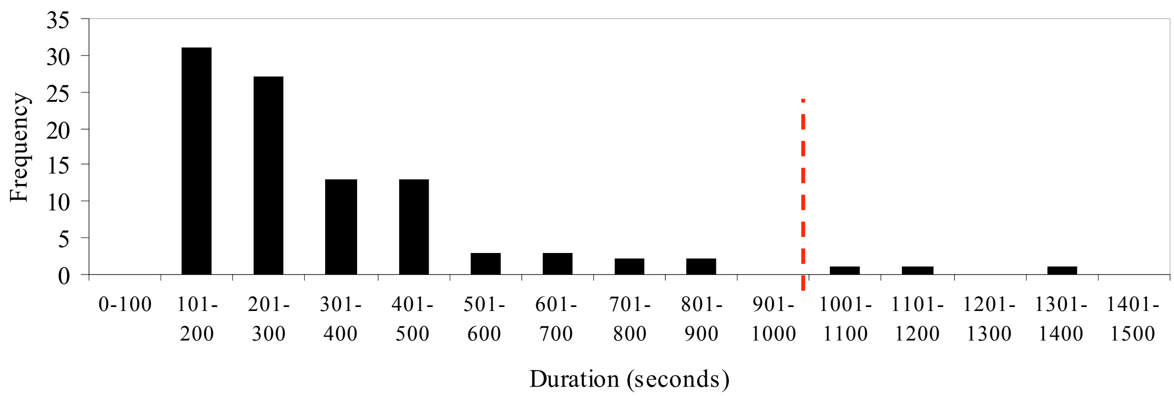


Figure E2. Duration distribution for the Vertical hitch task (*M* = 314.53, *SD* = 227.15).

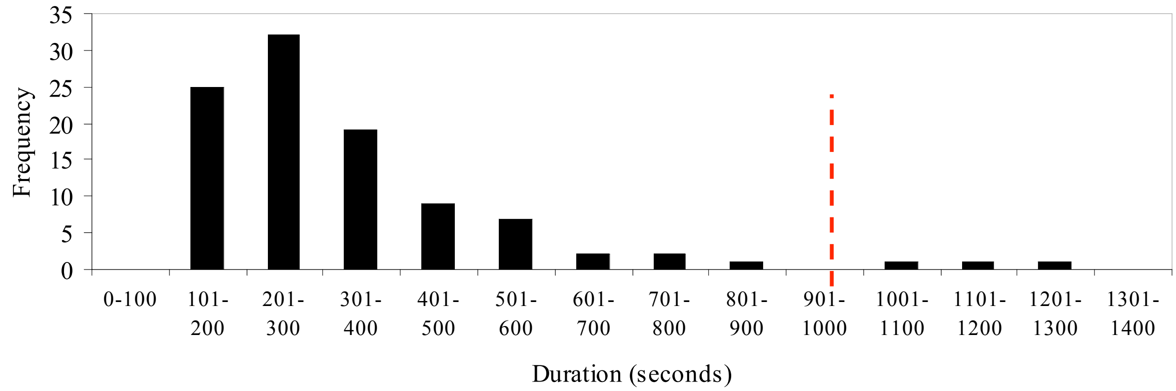


Figure E3. Duration distribution for the Pretzel knot task (*M* = 326.01, *SD* = 211.45).

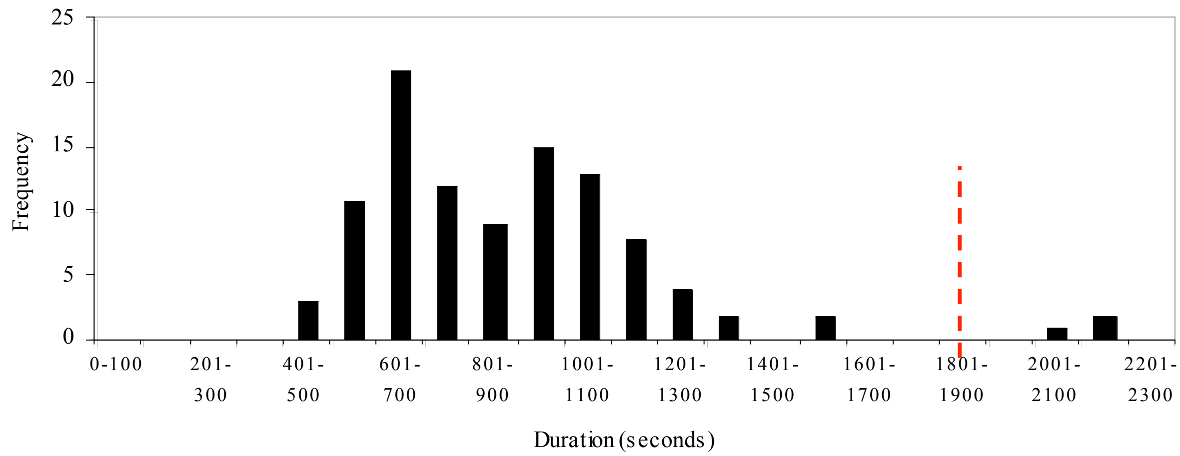


Figure E4. Duration distribution for the Crystal structure task (*M* = 896.97, *SD* = 316.65).

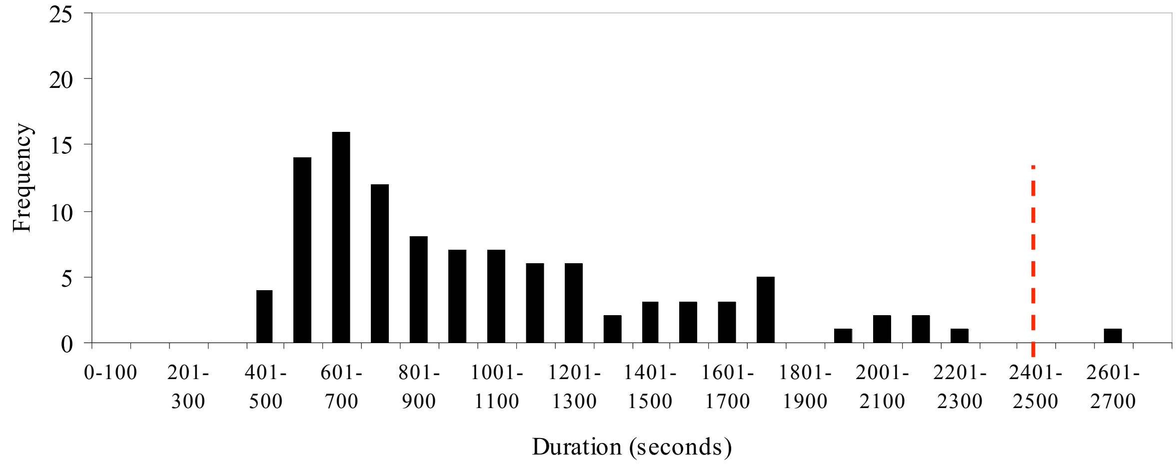


Figure E5. Duration distribution for the DNA structure task (*M* = 1070.10, *SD* = 475.59).

Isomorphic tasks (assessment session)

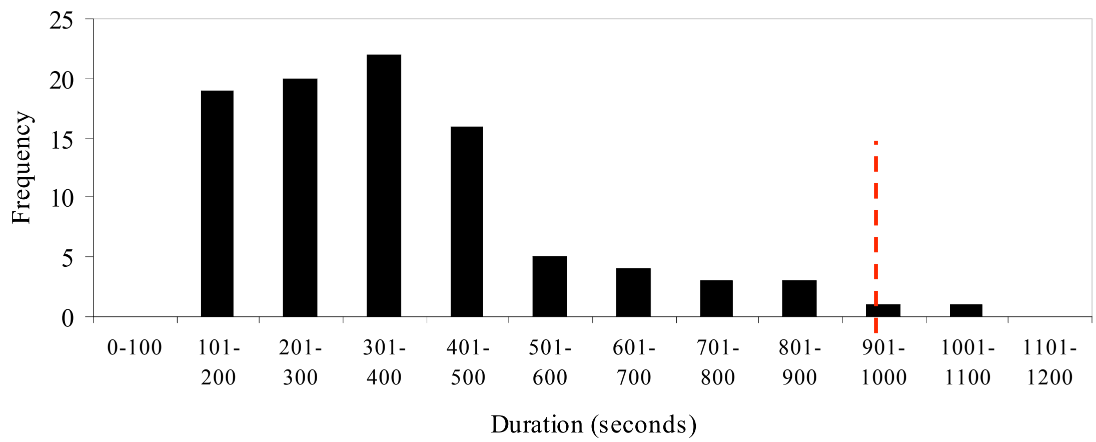


Figure E6. Duration distribution for the isomorphic Square knot task (*M* = 359.77, *SD* = 194.69).

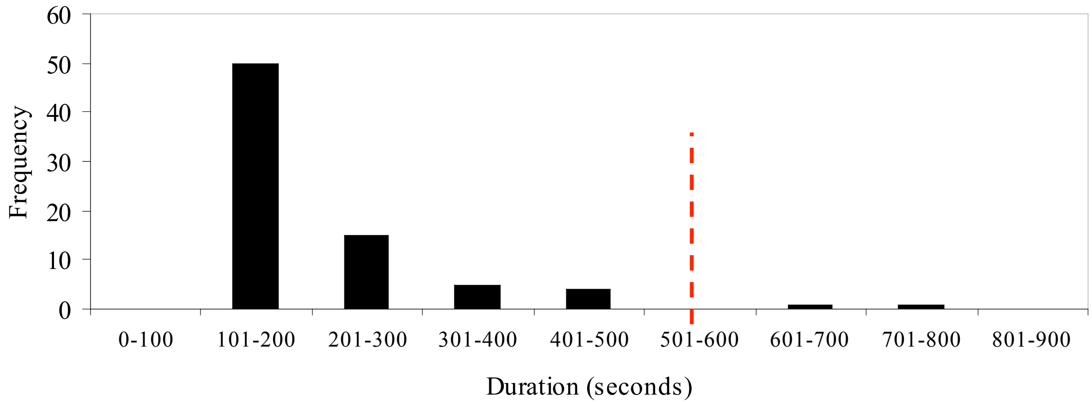


Figure E7. Duration distribution for the isomorphic Vertical hitch task (M = 183.05, SD = 121.61).

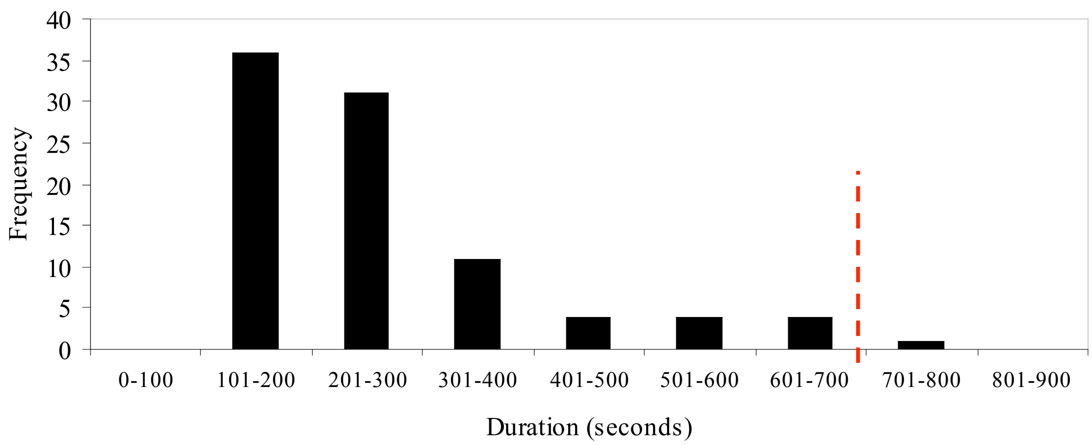


Figure E8. Duration distribution for the isomorphic Pretzel knot task (M = 258.02, SD = 146.49).

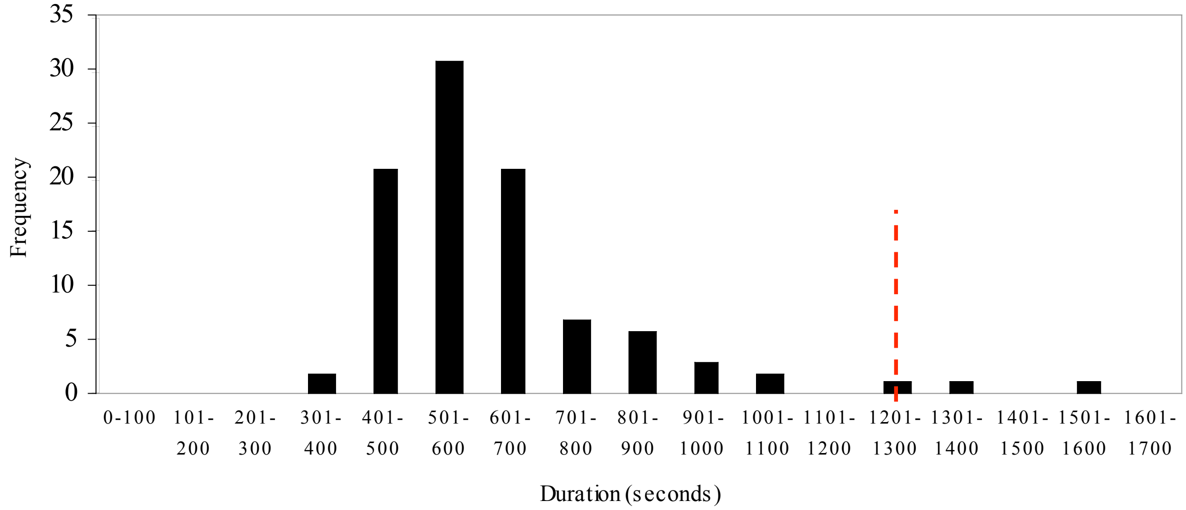


Figure E9. Duration distribution for the isomorphic Crystal structure task (M = 630.64, SD = 198.72).

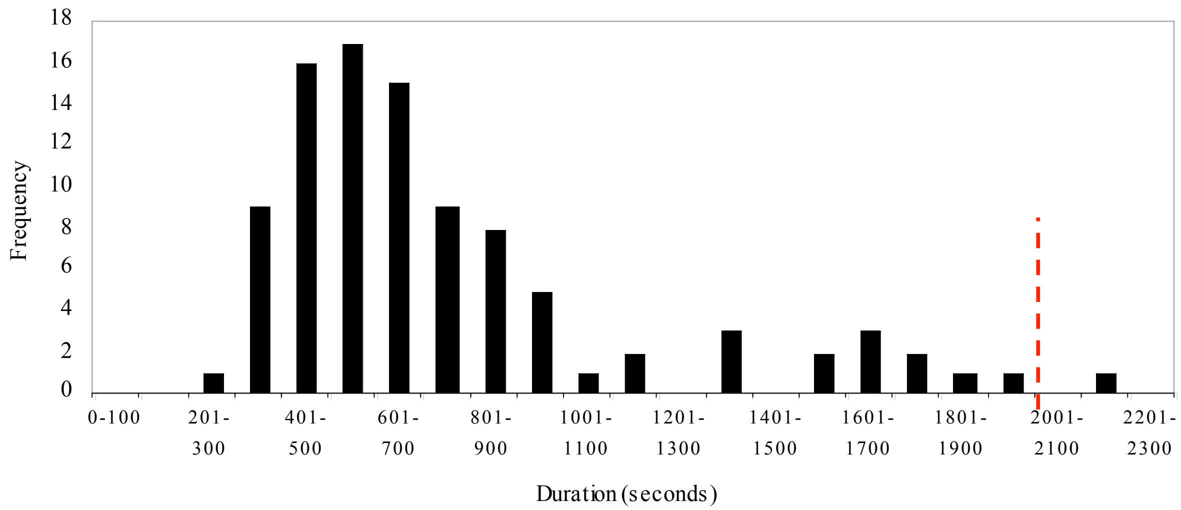


Figure E10. Duration distribution for the isomorphic DNA structure task (M = 757.89, SD = 415.76).

Novel tasks (training session)

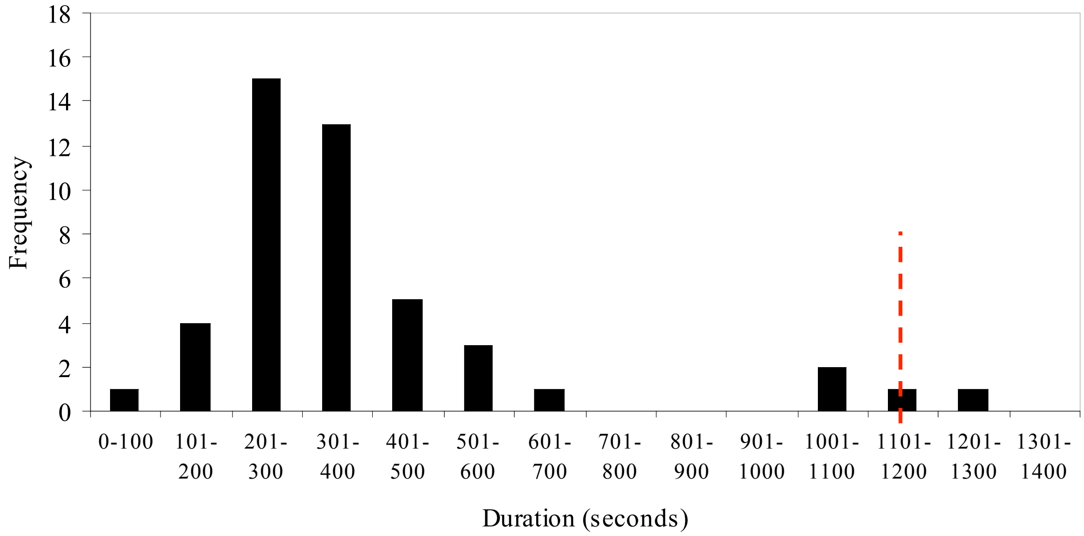


Figure E11. Duration distribution for the Square-twist knot task in the training session (*M* = 386.20, *SD* = 255.54).

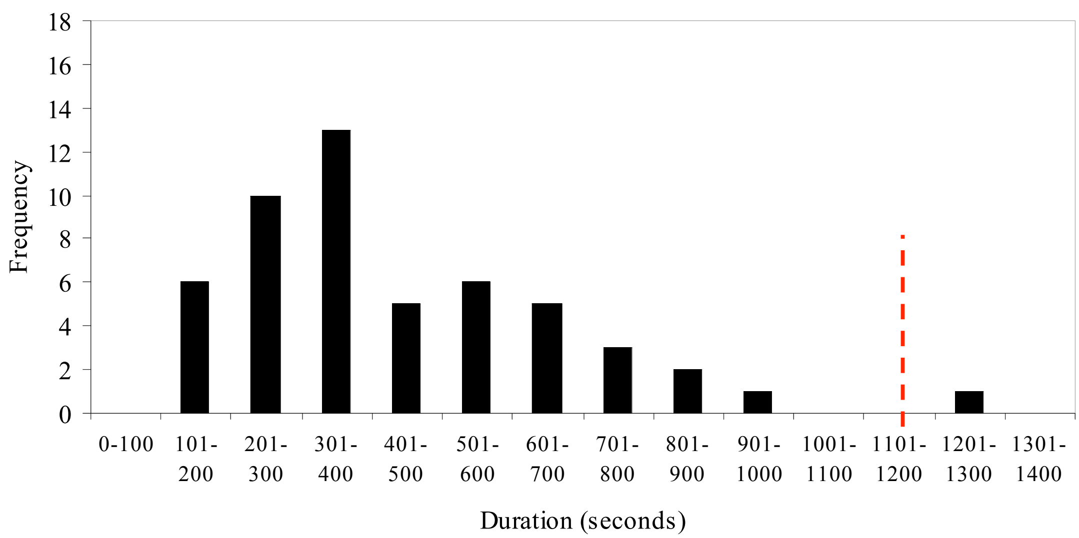


Figure E12. Duration distribution for the Hitch-chain task in the training session (*M* = 447.35, *SD* = 236.45).

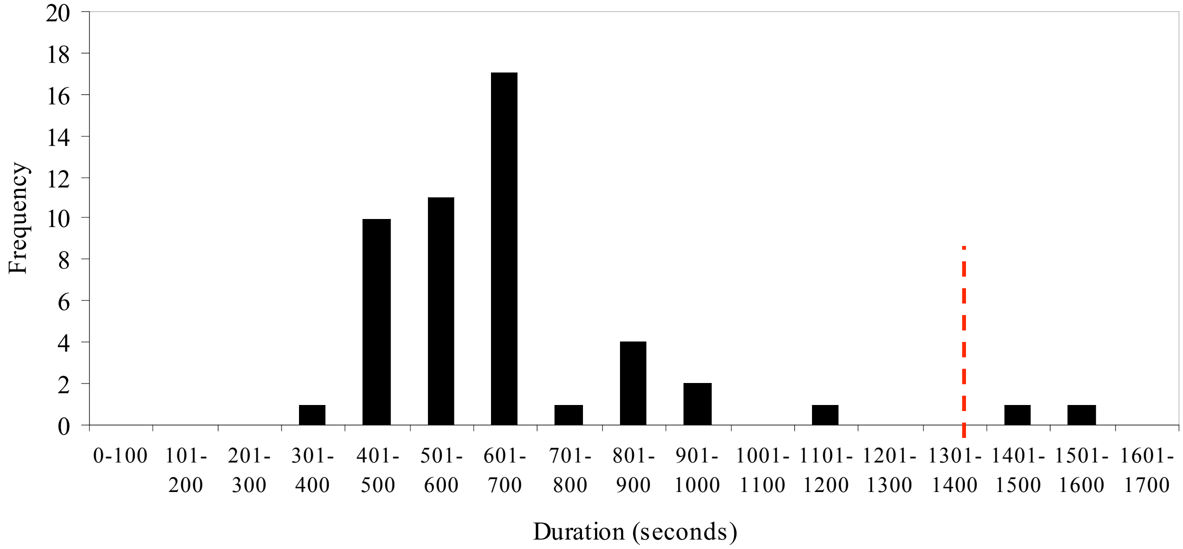


Figure E13. Duration distribution for the Element assembly task in the training session (M = 650.47, SD = 239.50).

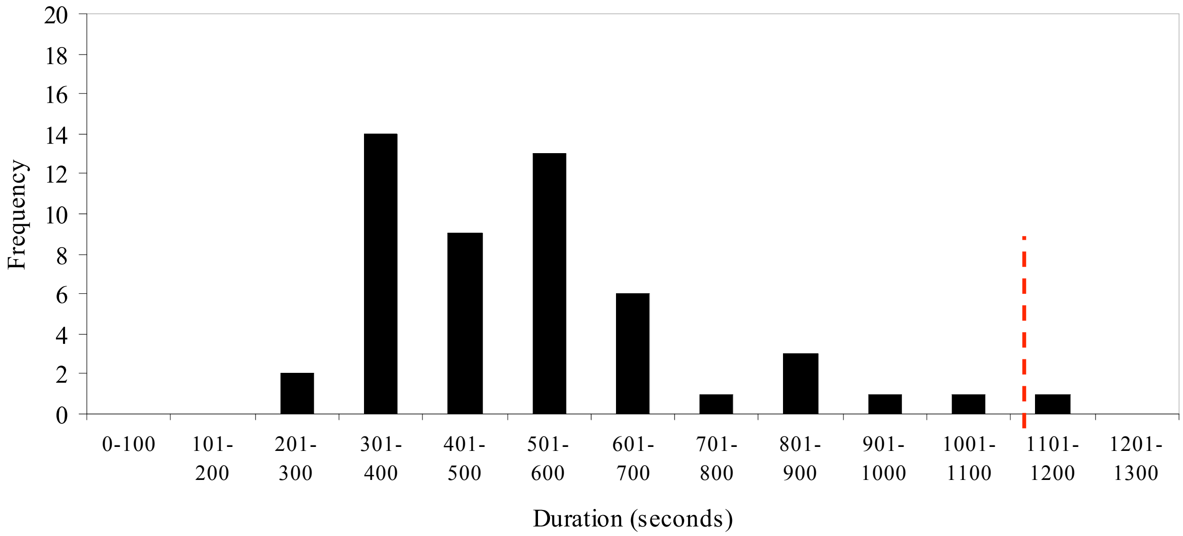


Figure E14. Duration distribution for the Cube assembly task in the training session (M = 529.61, SD = 194.15).

Novel tasks (assessment session)

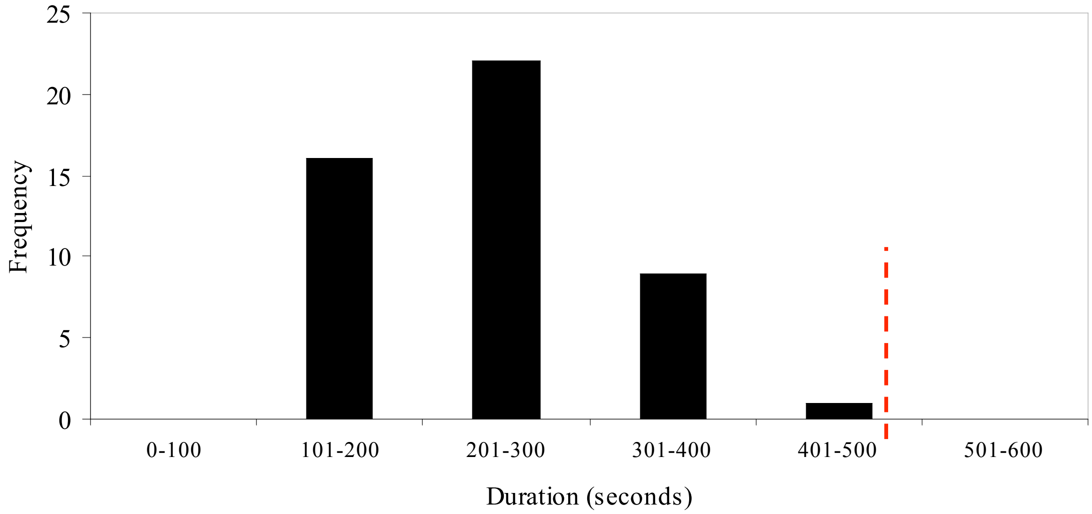


Figure E15. Duration distribution for the Square-twist knot task in the assessment session (*M* = 239.85, *SD* = 79.53).

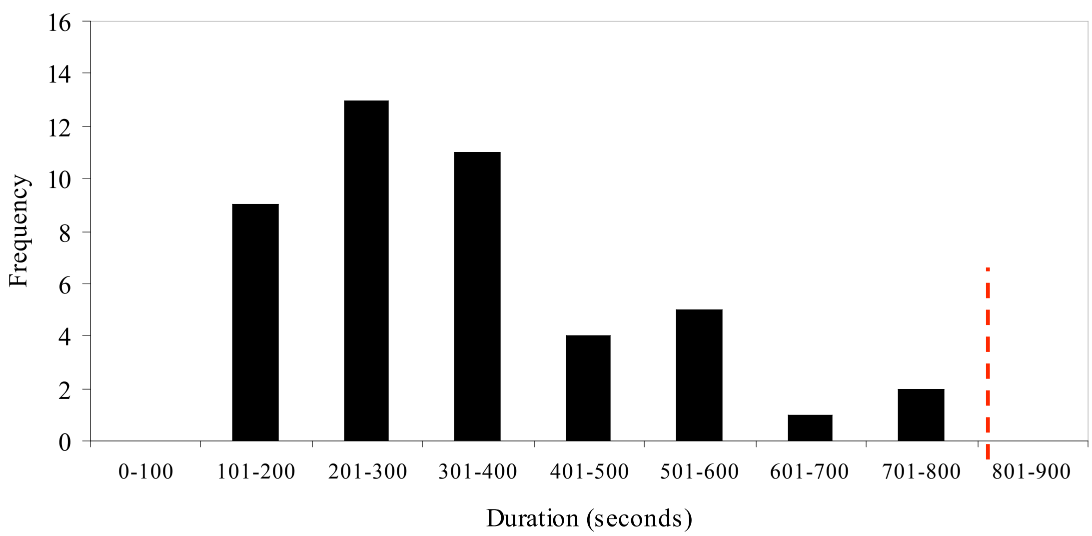


Figure E16. Duration distribution for the Hitch-chain task in the assessment session (*M* = 335.33, *SD* = 158.03).

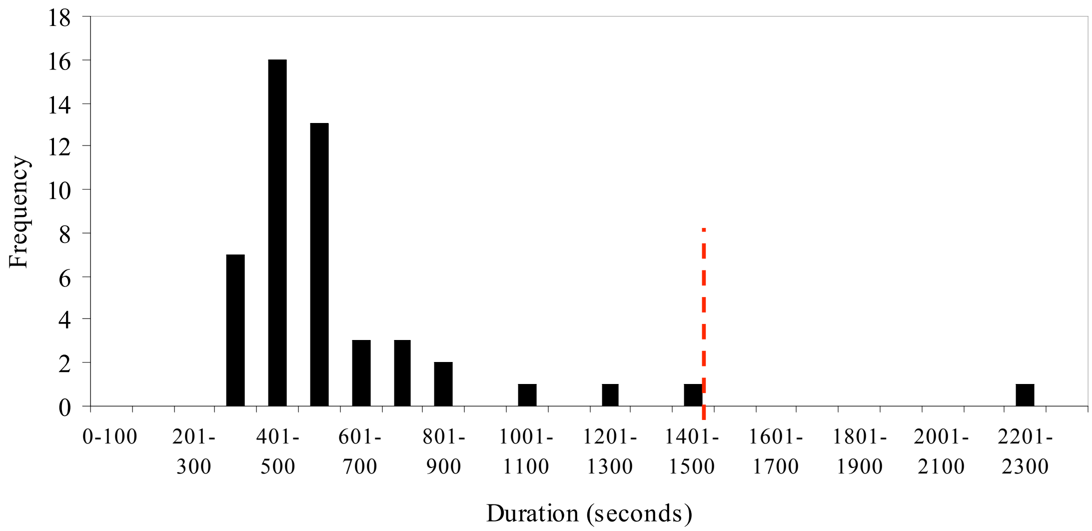


Figure E17. Duration distribution for the Element assembly task in the assessment session (M = 595.88, SD = 325.60).

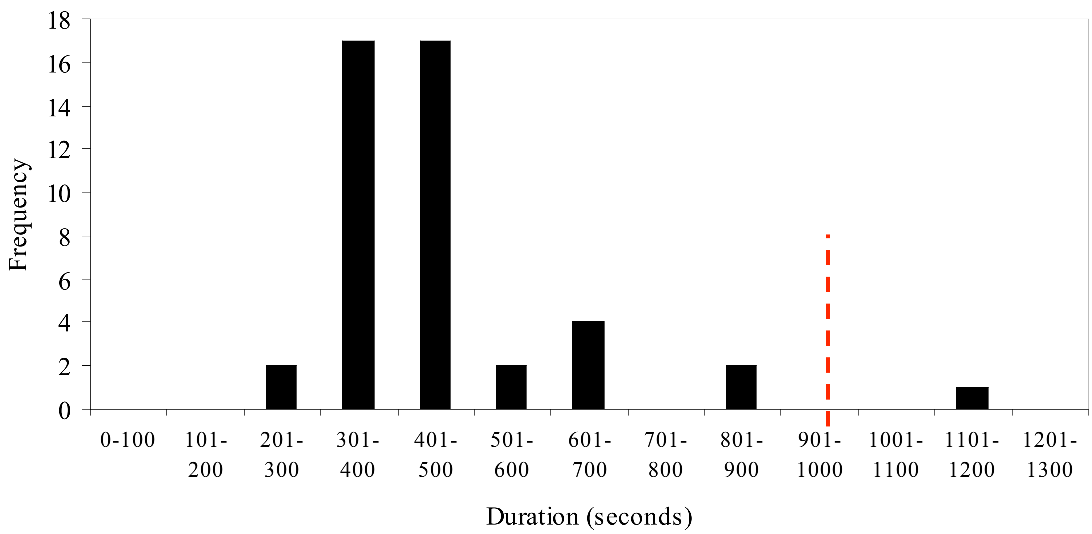


Figure E18. Duration distribution for the Cube assembly task in the assessment session (M = 456.62, SD = 168.86).

APPENDIX F

QUESTIONNAIRE RESULTS

Post-macramé task questionnaire

Table F1
Percentage of participants having done something similar to the knot tasks before.

Have you ever done anything like the knot tasks before?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Yes	46%	22%	35%	39%
No	54%	78%	65%	61%

Table F2
Participants' opinion (in percentages) as to what extent having done something similar to the macramé tasks before was helpful with the current tasks (only those participants saying they had done something similar before).

If you have ever done anything like the knot task before, did it help you with the current task?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	0%	0%	0%
Not really	0%	0%	10%	0%
Neutral	17%	0%	0%	0%
Somewhat	25%	60%	60%	60%
Very much	58%	40%	30%	40%

Table F3

Participants' opinion (in percentages) on how likeable they found the macramé tasks.

Did you like the knot tasks?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	4%	0%	16%
Not really	12%	22%	10%	8%
Neutral	23%	9%	0%	12%
Somewhat	39%	22%	35%	40%
Very much	27%	44%	55%	24%

Table F4

Participants' opinion (in percentages) on how difficult they found the macramé tasks.

Did you find the knot tasks difficult?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	12%	0%	7%	8%
Not really	31%	35%	38%	16%
Neutral	27%	9%	21%	12%
Somewhat	23%	44%	35%	40%
Very much	8%	13%	0%	24%

Table F5

Participants' opinion (in percentages) on how helpful they found the instructions for the macramé tasks.

Were the instructions helpful?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	4%	7%	8%
Not really	4%	9%	3%	16%
Neutral	4%	22%	14%	16%
Somewhat	35%	22%	41%	40%
Very much	58%	44%	35%	20%

Table F6

Participants' opinion (in percentages) on how usable they found the instructions for the macramé tasks.

Did you find the instructions easy to use?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	4%	7%	8%
Not really	0%	13%	3%	12%
Neutral	12%	17%	17%	12%
Somewhat	31%	17%	31%	36%
Very much	58%	48%	41%	32%

Table F7

Participants' report (in percentages) on which part of the instructions for the macramé tasks they focused most on.

When using the instructions, which part did you focus the most on?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
The pictures primarily	15%	55%	38%	71%
Mostly the pictures, but skimmed the text	39%	18%	35%	21%
The pictures and text equally	39%	23%	21%	4%
Mostly the text, but glanced at the pictures	8%	5%	7%	4%
The text primarily	0%	0%	0%	0%

Post-assembly task questionnaire

Table F8

Percentage of participants having done something similar to the assembly tasks before.

Have you ever used a construction kit like the one used to build the structures?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Yes	39%	39%	39%	44%
No	61%	61%	61%	56%

Table F9

Participants' opinion (in percentages) as to what extent having done something similar to the assembly tasks before was helpful with the current tasks (only those participants saying they had done something similar before).

If you have ever used a construction kit like the one used to build the structures before, did it help you with the current task?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	0%	9%	0%
Not really	40%	11%	18%	9%
Neutral	20%	22%	9%	18%
Somewhat	20%	44%	46%	64%
Very much	20%	22%	18%	9%

Table F10

Participants' opinion (in percentages) on how likeable they found the assembly tasks.

Did you like the structure tasks?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	0%	0%	4%
Not really	4%	17%	4%	16%
Neutral	27%	9%	4%	4%
Somewhat	31%	26%	39%	20%
Very much	39%	48%	54%	56%

Table F11

Participants' opinion (in percentages) on how difficult they found the assembly tasks.

Did you find the structure tasks difficult?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	12%	13%	7%	28%
Not really	27%	30%	50%	24%
Neutral	27%	13%	29%	12%
Somewhat	31%	44%	14%	32%
Very much	4%	0%	0%	4%

Table F12

Participants' opinion (in percentages) on how helpful they found the instructions for the assembly tasks.

Were the instructions helpful?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	4%	14%	4%	8%
Not really	23%	14%	19%	24%
Neutral	19%	41%	37%	20%
Somewhat	35%	18%	33%	32%
Very much	19%	14%	7%	16%

Table F13

Participants' opinion (in percentages) on how usable they found the instructions for the assembly tasks.

Did you find the instructions easy to use?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not at all	0%	9%	0%	8%
Not really	12%	9%	7%	4%
Neutral	15%	50%	30%	28%
Somewhat	35%	18%	41%	24%
Very much	39%	14%	22%	36%

Table F14

Participants' report (in percentages) on which part of the instructions for the assembly tasks they focused most on.

When using the instructions, which part did you focus the most on?	Inst-Before	Inst-Later	Inst-Simul	No-Inst
The pictures primarily	20%	40%	44%	63%
Mostly the pictures, but skimmed the text	36%	25%	30%	25%
The pictures and text equally	32%	15%	19%	13%
Mostly the text, but glanced at the pictures	12%	15%	7%	0%
The text primarily	0%	5%	0%	0%

Final questionnaire

Table F15

Participants' report (in percentages) on how they usually use instructions for procedural tasks.

How do you usually use instructions? (For example instructions for using, assembling, or troubleshooting things).	Inst-Before	Inst-Later	Inst-Simul	No-Inst
I read the instructions thoroughly before beginning the task	12%	4%	0%	12%
I skim the instructions shortly before beginning the task	27%	17%	39%	36%
I reference the instructions once in a while to verify what I am doing	31%	35%	32%	40%
I only use instructions if I run into problems	23%	44%	29%	4%
I never use instructions	8%	0%	0%	8%

Table F16

Participants' familiarity (in percentages) with using Lego.

Please indicate your familiarity with using Lego	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not familiar	8%	9%	4%	0%
Somewhat familiar	23%	17%	29%	32%
Very familiar	69%	74%	68%	68%

Table F17

Participants' familiarity (in percentages) with using Magnetix.

Please indicate your familiarity with using Magnetix	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not familiar	54%	70%	64%	64%
Somewhat familiar	39%	26%	32%	24%
Very familiar	8%	4%	4%	12%

Table F18

Participants' familiarity (in percentages) with using Mekano.

Please indicate your familiarity with using Mekano	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not familiar	89%	96%	96%	96%
Somewhat familiar	4%	0%	4%	0%
Very familiar	8%	4%	0%	4%

Table F19

Participants' familiarity (in percentages) with making friendship bracelets.

Please indicate your familiarity with making friendship bracelets	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not familiar	42%	44%	54%	60%
Somewhat familiar	23%	52%	36%	24%
Very familiar	35%	4%	11%	16%

Table F20

Participants' familiarity (in percentages) with braiding.

Please indicate your familiarity with braiding	Inst-Before	Inst-Later	Inst-Simul	No-Inst
Not familiar	28%	30%	32%	48%
Somewhat familiar	32%	48%	46%	32%
Very familiar	40%	22%	21%	20%

REFERENCES

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*(1), 41 – 47.

Alterman, R., Zito-Wolf, R., & Carpenter, T. (1991). Interaction, comprehension, and instruction usage. *The Journal of the Learning Sciences, 1*(3&4), 361 – 398.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35 – 44.

Carroll, J. M., Mack, R. L., Lewis, C. H., Grischkowsky, N. L., & Robertson, S. R. (1985). Exploring exploring a word processor. *Human-Computer Interaction, 1*, 283 – 307.

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self explanations improves understanding. *Cognitive Science, 18*, 439 – 477.

Duff, S. C., & Barnard, P. J. (1990). Influencing behavior via device representation: Decreasing performance by increasing instruction. *Human-Computer Interaction – INTERACT '90,* 61- 66.

Ekstrom, R. B, French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*(2), 3 – 84.

Fischer, G., & Sugimoto, M. (2006). Supporting Self-Directed Learners and Learning Communities with Sociotechnical Environments. *International Journal Research and Practice in Technology Enhanced Learning*, *1*(1), 31 – 64

Ganier (2004). Factors affecting the processing of procedural instructions: Implications for document design. *IEEE Transactions on Professional Communications, 47*(1), 15-26.

Green, A. J. K. (2002). Learning procedures and goal specificity in learning and problem-solving tasks. *European Journal of Cognitive Psychology, 14*(1), 105 – 126.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Ed.), Human mental workload, (pp. 139-183). Amsterdam, Netherlands: North-Holland.

Hickman, J. M, Rogers, W. A., & Fisk, A. D. (in press). Training older adults to use new technology. *Journal of Gerontology: Psychological Sciences*.

Kalyuga, S., Chandler, P., Tuovinen, J. & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579-588.

Koubek, R. J., Benysh, S. A. H. & Tang, E. (1997). Learning. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (Second edition), (pp.130-149). New York, NY: John Wiley & Sons.

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 4, pp. 181 – 248). Hillsdale, New Jersey: Lawerence Erlbaum Associates.

Lunger, M. (1998). *Hemp masters: Ancient hippie secrets for knotting hip hemp jewelry*. Liberty, UT: Eagle's View Publishing.

Marcus, N., Cooper, M. & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*(1), 49-63.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge, England: Cambridge University Press.

Meilach, D. Z. (1971). *Macramé: Creative Design in Knotting*. New York: Crown Publishers.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1988). *The design of everyday things*. New York: Basic Books.

Ohlsson, S. (1996). Learning from performance errors. *Psychological Review, 103*, 241 – 262.

Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510 – 532.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1 – 29.

Rogers, W. A. (1997). Individual differences, aging, and human factors: An overview. In A. D. Fisk & W. A. Rogers (Eds.), Handbook of human factors and the older adult. San Diego, CA: Academic Press.

Schriver, K. A. (1997). *Dynamics in document design*. Canada: John Wiley & Sons, Inc.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251 – 296.

Tuovinen, J. E. & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334-341.

Vakil, E., Hoffman, Y., & Myzliek, D. (1998). Active versus passive procedural learning in older and younger adults. *Neuropsychological Rehabilitation, 8*(1), 31 – 41.

Zometool, Inc (2006). *Zome Kit Brochures*. Retrieved March 8, 2006, from http://www.zometool.com/build-more.html